R-Estimation for Asymmetric Independent Component Analysis

Marc Hallin* and Chintan Mehta†

April 11, 2013

Abstract

Independent Component Analysis (ICA) recently has attracted much attention in the statistical literature as an attractive and useful alternative to elliptical models. Whereas k-dimensional elliptical densities depend on one single unspecified radial density, however, k-dimensional independent component distributions involve k unspecified component densities. In practice, for a given sample size n and given dimension k, this makes the statistical analysis much harder. We focus here on the estimation, from an independent sample, of the mixing/demixing matrix of the model. Traditional methods (FOBI, Kernel-ICA, FastICA) mainly originate from the engineering literature. The statistical properties of those methods are not well known, and they typically require very large samples. So does the "classical semiparametric" approach by Chen and Bickel (2006), which is based on an estimation of the k component densities (those densities being those of the unobserved independent components). The "double scatter matrix" method of Oja et al. (2006) and (2008) requires the arbitrary choice of two scatter matrices generally based on estimated higher-order moments which are likely to be poorly robust. As a reaction, an efficient (signed-)rank-based approach has been proposed by Ilmonen and Paindaveine (2011) for the case of symmetric component densities; their estimators unfortunately fail to be root-n consistent as soon as one of the component densities violates the symmetry assumption. In this paper, using ranks rather than signed ranks, we extend their approach to the asymmetric case and propose a one-step R-estimator for ICA mixing matrices. The finite-sample performances of those estimators are investigated and compared to those of existing methods under moderately large sample sizes. Particularly good performances are obtained from a version involving data-driven scores taking into account the skewness and kurtosis of residuals. Finally, we show, by an empirical exercise, that our methods also may provide excellent results in contexts such as image analysis, where the basic assumptions of ICA are quite unlikely to hold.

Keywords and phrases: Independent Component Analysis (ICA), local asymptotic normality (LAN), ranks, R-estimation, robustness.

*Marc Hallin is Professor of Statistics, ECARES, Université libre de Bruxelles, Ave. F.D. Roosevelt 50, CP 114/04, B-1050 Bruxelles, Belgium and ORFE, Princeton University, Sherrerd Hall, Princeton, NJ 08544 (E-mail: mhallin@ulb.ac.be). Research supported by the Sonderforschungsbereich "Statistical modelling of nonlinear dynamic processes" (SFB 823) of the Deutsche Forschungsgemeinschaft, and the IAP research network grant P7/06 of the Belgian Federal Government (Belgian Science Policy).

†Chintan Mehta is Doctoral Student at ORFE, Princeton University, Sherrerd Hall, Princeton, NJ 08544 (E-mail: cmehta@princeton.edu). The research was partially supported by the National Institute of General Medical Sciences of the National Institutes of Health through Grant Number R01-GM072611 and National Science Foundation grant DMS-0739195.

1 Introduction

1.1 Independent Component Analysis (ICA)

The traditional Gaussian model for noise, where a k-dimensional error term $\mathbf{e}$ is $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ can be extended, mainly, into two directions. Either the elliptical density contours of the multinormal are preserved, and $\mathbf{e}$ is assumed to be *elliptically symmetric* with respect to the origin, with unspecified radial density f. Or, the independence of the marginals of $\mathbf{\Sigma}^{-1/2}\mathbf{e}$ is preserved, but their densities $f_1, \ldots, f_k$ remain unspecified, yielding the *independent component* model. In both cases, the distribution of $\mathbf{e}$ involves an unknown linear transformation—the $k \times k$ symmetric positive definite *sphericizing matrix* $\mathbf{\Sigma}^{-1/2}$, that is, $k(k+1)/2$ parameters, in the elliptical case, the $k \times k$ *mixing matrix* $\mathbf{\Lambda}$ (equivalently, the *demixing matrix* $\mathbf{\Lambda}^{-1}$, that is, k^2 parameters in the independent component case. The main difference, however, is that, while elliptical noise only depends on one nonparametric nuisance, the radial density f, independent component noise involves k nonparametric nuisances, the component densities $f_1, \ldots, f_k$. This makes the statistical analysis of models based on independent component noise significantly harder than its elliptical counterpart: for given k and n, for instance, estimating $\mathbf{\Lambda}$ is much more difficult than estimating $\mathbf{\Sigma}$.

In this paper, we focus on the problem of estimating $\mathbf{\Lambda}$. Many solutions—FastICA, Kernel-ICA, FOBI, ... have been proposed, mostly in the engineering literature; see Section 4.1 for details. They typically require very large samples, and their statistical properties are not always well known. A method based on the availability of two scatter matrices has been developed by Oja et al. (2006) and (2008), and involves the somewhat arbitrary choice of two scatter matrices, generally based on estimated higher-order moments which are likely to be poorly robust and quite sensitive to possibly heavy tails in some of the component densities; also, a symmetrization step is required by the method, which is computationally quite demanding. A rigorous asymptotic analysis of the problem is provided by Chen and

Bickel (2006) in line with Bickel at al. (1993) 's classical semiparametric methodology, based on tangent space projections. In that approach, the k component densities $f_1, \ldots, f_k$ need to be estimated, which again is very costly, and requires very large sample sizes.

As a reaction, an efficient rank-based method has been developed recently by Ilmonen and Paindaveine (2011), taking into account the invariance and distribution-freeness features of ranks in order to bypass the costly step of estimating k densities. The performance of their estimators—call them R_+-estimators—is quite good, even under moderately large samples. However, they are based on marginal *signed ranks*, which requires the somewhat restrictive assumption that all component densities are symmetric.

We show how that unpleasant assumption can be avoided, and propose a one-step R-estimation procedure based on residual ranks rather than the residual signed ranks used in R_+-estimation. We establish the asymptotic root-n consistency and asymptotic normality of our R-estimators, and carefully study their finite-sample performances via simulations. In particular, we show how they improve on the traditional methods (FOBI, FastICA, Kernel-ICA, and some others), and outperform Ilmonen and Paindaveine's R_+-estimators as soon as the symmetry assumption is violated (in which case their estimators are no longer root-n consistent). R-estimation, as well as R_+-estimation, in this context, requires choosing k score functions, a choice that in practice may be somewhat difficulty. We therefore describe and recommend a version of our method based on data-driven scores, where the skewness and kurtosis of component residuals are taken into account. That method is easily implementable, and achieves particularly good results.

Finally, with an application to image analysis, we also show that our method also provides good results in situations where the basic assumptions of ICA clearly do not hold. There, our R-estimators are shown to improve, quite substantially, the demixing performances of such classical methods as FOBI, FastICA or Kernel-ICA.

1.2 Notation, identifiability and main assumptions

Denote by $\mathbf{X}^{(n)} := (\mathbf{X}_1^{(n)\prime}, \ldots, \mathbf{X}_n^{(n)\prime})'$, $n \in \mathbb{N}$, with $\mathbf{X}_i^{(n)\prime} := (X_{i1}^{(n)}, \ldots, X_{ik}^{(n)})$, $i = 1, \ldots, n$, a triangular array of observed k-dimensional random vectors satisfying

$$\mathbf{X}_i^{(n)} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{Z}_i^{(n)} \tag{1.1}$$

where $\mathbf{Z}^{(n)} := (\mathbf{Z}_1^{(n)\prime}, \ldots, \mathbf{Z}_n^{(n)\prime})'$ is an unobserved n-tuple of i.i.d. k-dimensional *latent vectors* $\mathbf{Z}_i^{(n)\prime} := (Z_{i1}^{(n)}, \ldots, Z_{ik}^{(n)})$, $i = 1, \ldots, n$, with joint and marginal densities $f^{\mathbf{Z}}$ and $f_1, \ldots, f_k$ such that

$$f^{\mathbf{Z}}(\mathbf{z}) = \prod_{j=1}^{k} f_j(z_j), \qquad \mathbf{z} = (z_1, \ldots, z_k) \in \mathbb{R}^k. \tag{1.2}$$

The $k \times 1$ vector $\boldsymbol{\mu}$ and the $k \times k$ full-rank matrix $\boldsymbol{\Lambda}$ are parameters; $\boldsymbol{\Lambda}$ and its inverse $\boldsymbol{\Lambda}^{-1}$ are called the *mixing* and *demixing* (or *unmixing*) matrices, respectively. Under (1.2), the k components $Z_{i1}^{(n)}, \ldots, Z_{ik}^{(n)}$ of the latent vectors $\mathbf{Z}_i^{(n)}$ are mutually independent: they are called the *independent components*, and their marginal probability densities $f := (f_1, \ldots, f_k)$ the *component densities*, of the *independent component model* (1.1)-(1.2).

Identification constraints clearly are needed in order for $\boldsymbol{\mu}$ and $\boldsymbol{\Lambda}$ to be identified. Without any loss of generality, we throughout impose that $f \in \mathcal{F}_0$, where

$$\mathcal{F}_0 := \Big\{ f := (f_1, \ldots, f_k) \mid f_j(z) > 0 \text{ for all } z \in \mathbb{R}, \text{ and } \int_{-\infty}^{0} f_j(z)\mathrm{d}z = 1/2 = \int_{0}^{\infty} f_j(z)\mathrm{d}z \Big\};$$

the vector $\boldsymbol{\Lambda}^{-1}\boldsymbol{\mu}$ then is identified as the componentwise median of the $\boldsymbol{\Lambda}^{-1}\mathbf{X}_i^{(n)}$'s. Identification issues for $\boldsymbol{\Lambda}$ are more severe, due to the invariance of the IC assumption (1.1) and (1.2) under permutation, rescaling, and sign changes of the centered independent components $\mathbf{Z}_i^{(n)} - \boldsymbol{\Lambda}^{-1}\boldsymbol{\mu}$. Denoting by $\mathbf{D}_1$ and $\mathbf{D}_2$ two arbitrary full-rank $k \times k$ diagonal matrices, and by $\mathbf{P}$ an arbitrary $k \times k$ permutation matrix, we clearly have that $\boldsymbol{\Lambda}\mathbf{Z} = \boldsymbol{\Lambda}^*\mathbf{Z}^*$ for $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\mathbf{D}_1\mathbf{P}\mathbf{D}_2$ and $\mathbf{Z}^* = \mathbf{D}_2^{-1}\mathbf{P}^{-1}\mathbf{D}_1^{-1}\mathbf{Z}$, where $\mathbf{Z}^*$ still satisfies (1.1) and (1.2). The

mixing matrices $\mathbf{\Lambda}$ and $\mathbf{\Lambda}^*$ therefore are observationally equivalent.

Several identification constraints have been proposed in the literature in order to tackle this identifiability issue. Those we are imposing here are borrowed from Ilmonen and Paindaveine (2011). Considering the equivalence classes of $k \times k$ nonsingular matrices associated with the equivalence relation $\mathbf{\Lambda}^* \sim \mathbf{\Lambda}$ iff $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{D}_1\mathbf{P}\mathbf{D}_2$ for some permutation and full-rank diagonal matrices $\mathbf{P}$, $\mathbf{D}_1$ and $\mathbf{D}_2$, respectively, denote by Π the mapping

$$\mathbf{\Lambda} \mapsto \Pi(\mathbf{\Lambda}) := \mathbf{\Lambda}\mathbf{D}_1^{\mathbf{\Lambda}}\mathbf{P}^{\mathbf{\Lambda}}\mathbf{D}_2^{\mathbf{\Lambda}}, \tag{1.3}$$

where (a) $\mathbf{D}_1^{\mathbf{\Lambda}}$ is the $k \times k$ positive diagonal matrix whose j^{th} diagonal element is the inverse of the Euclidean norm of $\mathbf{\Lambda}$'s j^{th} column ($j = 1, \ldots, k$), (b) $\mathbf{P}^{\mathbf{\Lambda}}$ is a permutation matrix that reorders the columns of $\mathbf{\Lambda}\mathbf{D}_1^{\mathbf{\Lambda}}$ in such a way that $\left|(\mathbf{\Lambda}\mathbf{D}_1^{\mathbf{\Lambda}}\mathbf{P}^{\mathbf{\Lambda}})_{ij}\right| < \left|(\mathbf{\Lambda}\mathbf{D}_1^{\mathbf{\Lambda}}\mathbf{P}^{\mathbf{\Lambda}})_{ii}\right|$ for all $j > i$, and (c) the (not necessarily positive) diagonal matrix $\mathbf{D}_2^{\mathbf{\Lambda}}$ normalizes $\mathbf{\Lambda}\mathbf{D}_1^{\mathbf{\Lambda}}\mathbf{P}^{\mathbf{\Lambda}}$ in such a way that $(\mathbf{\Lambda}\mathbf{D}_1^{\mathbf{\Lambda}}\mathbf{P}^{\mathbf{\Lambda}}\mathbf{D}_2^{\mathbf{\Lambda}})_{jj} = 1$, i.e. $(\mathbf{D}_2^{\mathbf{\Lambda}})_{jj} = (\mathbf{\Lambda}\mathbf{D}_1^{\mathbf{\Lambda}}\mathbf{P}^{\mathbf{\Lambda}})_{jj}^{-1}$ for $j = 1, \ldots, k$. Consider the set $\mathcal{M}_k$ of nonsingular $k \times k$ matrices for which no tie occurs in the definition of $\mathbf{P}^{\mathbf{\Lambda}}$. Then, for $\mathbf{\Lambda}_1, \mathbf{\Lambda}_2 \in \mathcal{M}_k$, $\mathbf{\Lambda}_1 \sim \mathbf{\Lambda}_2$ if and only if $\Pi(\mathbf{\Lambda}_1) = \Pi(\mathbf{\Lambda}_2)$. Each class of equivalence thus contains a unique element $\mathbf{\Lambda}$ such that $\Pi(\mathbf{\Lambda}) = \mathbf{\Lambda}$, and inference for mixing matrices can be restricted to the set $\mathcal{M}_k^1 := \Pi(\mathcal{M}_k)$.

The matrices $\mathbf{\Lambda}$ for which ties occur in the construction of $\mathbf{P}^{\mathbf{\Lambda}}$ have Lebesgue measure zero in $\mathbb{R}^{k\times k}$; neglecting them has little practical implications. While one could devise a systematic way to define a unique $\mathbf{P}^{\mathbf{\Lambda}}$ in the presence of such ties, the resulting mapping $\mathbf{\Lambda} \mapsto \mathbf{P}^{\mathbf{\Lambda}}$ would not be continuous, which disallows the use of the Delta method when constructing root-n consistent estimators for $\mathbf{\Lambda}$.

For $\mathbf{L} \in \mathcal{M}_k^1$, denote by $\boldsymbol{\theta} = (\boldsymbol{\mu}, \text{vecd}^\circ(\mathbf{L}))$ the model parameter, where $\text{vecd}^\circ(\mathbf{L})$ stands for the vector of size $k(k-1)$ that stacks the columns of $\mathbf{L}$ on top of each with the diagonal elements omitted (since, by definition, they are set to one). Write $\Theta := \left(\mathbb{R}^k \times \text{vecd}^\circ(\mathcal{M}_k^1)\right)$ for the parameter space. Note that, by imposing scaling and some nonnegative asymmetry

constraints on the component densities, one could add the (unique) diagonal matrix $\mathbf{D}^{\mathbf{\Lambda}}$ such that $\mathbf{D}_1^{\mathbf{\Lambda}}\mathbf{P}^{\mathbf{\Lambda}}\mathbf{D}_2^{\mathbf{\Lambda}} = \mathbf{P}^{\mathbf{\Lambda}}\mathbf{D}^{\mathbf{\Lambda}}$ to the list of (nuisance) parameters. In the present context, it is more convenient to have it absorbed into the unspecified form of f. The role of $\mathbf{D}^{\mathbf{\Lambda}}$ is quite similar, in that respect, to that of the scale functional in elliptical families, as discussed in Hallin and Paindaveine (2006).

Another solution to those identification problems is adopted by Chen and Bickel (2006), who impose scaling restrictions of f, and then let their PCFICA algorithm (Chen and Bickel 2005) make a choice between the various observationally equivalent values of the demixing matrix $\mathbf{\Lambda}^{-1}$. Specifically, they restrict to a parameter space for demixing matrices consisting of full-rank $k \times k$ matrices that satisfy the following: every row has unit norm, the element with largest absolute value in each row is positive, and rows are ordered by their maximum element. This parameter space, like $\mathcal{M}_k^1$, contains unique representatives from equivalence classes amongst observationally equivalent $k \times k$ full-rank matrices.

2 Local asymptotic normality and group invariance

2.1 Group Invariance and semiparametric efficiency

Denoting by $\mathrm{P}_{\boldsymbol{\theta};f}^{(n)}$, $\mathrm{P}_{\boldsymbol{\mu},\mathbf{L};f}^{(n)}$ or $\mathrm{P}_{\boldsymbol{\mu},\mathrm{vecd}^\circ(\mathbf{L});f}^{(n)}$ the joint distribution of $\mathbf{X}^{(n)}$ under location $\boldsymbol{\mu}$, mixing matrix $\mathbf{\Lambda}$ such that $\Pi(\mathbf{\Lambda}) = \mathbf{L}$, and component densities $f = (f_1, \ldots, f_k)$, let

$$\mathcal{P}^{(n)} := \left\{\mathrm{P}_{\boldsymbol{\theta};f}^{(n)} \mid \boldsymbol{\theta} \in \Theta,\ f \in \mathcal{F}_0\right\}; \qquad \mathcal{P}_f^{(n)} := \left\{\mathrm{P}_{\boldsymbol{\theta};f}^{(n)} \mid \boldsymbol{\theta} \in \Theta\right\} \text{ for fixed } f \in \mathcal{F}_0;$$

$$\mathcal{P}_{\boldsymbol{\mu};f}^{(n)} := \left\{\mathrm{P}_{\boldsymbol{\mu},\mathbf{L};f}^{(n)} \mid \mathbf{L} \in \mathcal{M}_k^1\right\} \text{ for fixed } \boldsymbol{\mu} \in \mathbb{R}^k \text{ and } f \in \mathcal{F}_0;$$

$$\mathcal{P}_{\mathbf{L}}^{(n)} \text{ or } \mathcal{P}_{\mathbf{\Lambda}}^{(n)} := \left\{\mathrm{P}_{\boldsymbol{\mu},\mathbf{L};f}^{(n)} \mid \boldsymbol{\mu} \in \mathbb{R},\ f \in \mathcal{F}_0\right\} \text{ for fixed } \Pi(\mathbf{\Lambda}) = \mathbf{L} \in \mathcal{M}_k^1; \quad \text{and}$$

$$\mathcal{P}_{\boldsymbol{\mu},\mathbf{L}}^{(n)} \text{ or } \mathcal{P}_{\boldsymbol{\mu},\mathbf{\Lambda}}^{(n)} := \left\{\mathrm{P}_{\boldsymbol{\mu},\mathbf{L};f}^{(n)} \mid f \in \mathcal{F}_0\right\} \text{ for fixed } \boldsymbol{\mu} \in \mathbb{R}^k \text{ and } \Pi(\mathbf{\Lambda}) = \mathbf{L} \in \mathcal{M}_k^1.$$

All those subfamilies will play a role in the sequel.

A semiparametric (in the spirit of Bickel et al. (1993)) approach to Independent Component Analysis (ICA) and, more particularly, the estimation of $\mathbf{\Lambda}$, requires the *uniform local asymptotic normality* (ULAN) of $\mathcal{P}_f^{(n)}$ at any f satisfying adequate regularity assumptions: see Section 2.2. It is easy to see that ULAN of $\mathcal{P}_f^{(n)}$ (with parameters $\boldsymbol{\mu}$ and $\mathbf{L}$) implies that of $\mathcal{P}_{\boldsymbol{\mu};f}^{(n)}$ (with parameter $\mathbf{L}$) for any given $\boldsymbol{\mu} \in \mathbb{R}^k$.

The model we are interested in involves the family $\mathcal{P}^{(n)}$. Depending on the context, several distinct semiparametric approaches to ICA are possible: either both the location $\boldsymbol{\mu}$ and the mixing matrix $\mathbf{\Lambda}$ are parameters of interest with the density f being a nuisance; or the location $\boldsymbol{\mu}$ is a parameter of interest with nuisance $(\mathbf{\Lambda}, f)$; or the mixing matrix $\mathbf{\Lambda}$ (equivalently, $\mathbf{L}$) only is of interest and $(\boldsymbol{\mu}, f)$ is a nuisance. Hallin and Werker (2003) have shown that, under very general conditions, if the parametric submodels associated with fixed values of the nuisance are *uniformly locally asymptotically normal* (ULAN), while the submodels associated with fixed values of the parameter of interest are generated by groups of transformations, then semiparametrically efficient inference can be based on the maximal invariants of those groups.

In the present context, $\mathbf{\Lambda}$ is the parameter of interest, and $(\boldsymbol{\mu}, f)$ is the nuisance. Consider $f = (f_{*1}, \ldots, f_{*k})$, and assume that

(A1) f belongs to the subset $\mathcal{F}_{\text{ULAN}}$ of $\mathcal{F}_0$ such that the sequence of (parametric) subfamilies $\mathcal{P}_{\boldsymbol{\mu};f}^{(n)}$, with parameter $\mathbf{L}$, is ULAN, with *central sequence* $\mathbf{\Delta}_{\boldsymbol{\mu};f}^{(n)}(\mathbf{L})$ (actually, ULAN holds at any $(\boldsymbol{\mu}, f)$ iff it holds at $(\mathbf{0}, f)$), and

(A2) for all $\mathbf{L} \in \mathcal{M}_k^1$ and $n \in \mathbb{N}$, the (nonparametric) subfamily $\mathcal{P}_{\mathbf{L}}^{(n)}$ is generated by some group of transformations $\mathcal{G}^{(n)}(\mathbf{L}), \circ$ acting on the observation space $\mathbb{R}^{kn}$, with maximal invariant $\mathbf{R}^{(n)}(\mathbf{L})$.

By Hallin and Werker (2003), the semiparametric efficiency bounds (at $(\boldsymbol{\mu}, f)$, for the problem where $\mathbf{L}$ is the parameter of interest) can be achieved by basing inference on the maximal in-

variant $\mathbf{R}^{(n)}(\mathbf{L})$—more specifically, on the conditional expectation $\mathrm{E}_{\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L};f}}[\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\mu};f}(\mathbf{L})|\ \mathbf{R}^{(n)}(\mathbf{L})]$; since $\mathbf{R}^{(n)}(\mathbf{L})$ is invariant, that conditional expectation moreover is distribution-free under $\mathcal{P}^{(n)}_{\mathbf{L}}$ (hence, also under densities f that do not necessarily belong to $\mathcal{F}_{\text{ULAN}}$).

Section 2.2 establishes the ULAN property (A1) of $\mathcal{P}^{(n)}_{\boldsymbol{\mu};f}$ for any $\boldsymbol{\mu}$ and f satisfying some mild regularity assumptions. Let us show here that (A2) holds for any $\mathbf{L} \in \mathcal{M}^1_k$ and n, and that the maximal invariant is the vector $\mathbf{R}^{(n)}(\mathbf{L}) = (\mathbf{R}^{(n)\prime}_1(\mathbf{L}), \ldots, \mathbf{R}^{(n)\prime}_n(\mathbf{L}))'$, where $\mathbf{R}^{(n)}_i(\mathbf{L}) = (R^{(n)}_{i1}(\mathbf{L}), \ldots, R^{(n)}_{ik}(\mathbf{L}))'$ and $R^{(n)}_{ij}(\mathbf{L})$ is the rank of $(\mathbf{L}^{-1}\mathbf{X}^{(n)}_i)_j$ among $(\mathbf{L}^{-1}\mathbf{X}^{(n)}_1)_j, \ldots, (\mathbf{L}^{-1}\mathbf{X}^{(n)}_n)_j$—hence also, under $\mathcal{P}^{(n)}_{\mathbf{L}}$, letting

$$\mathbf{Z}^{(n)}_i(\boldsymbol{\mu}, \mathbf{L}) := \mathbf{L}^{-1}(\mathbf{X}^{(n)}_i - \boldsymbol{\mu}), \quad i = 1, \ldots, n, \tag{2.4}$$

the rank of $\big(\mathbf{Z}^{(n)}_i(\boldsymbol{\mu}, \mathbf{L})\big)_j$ among $\big(\mathbf{Z}^{(n)}_1(\boldsymbol{\mu}, \mathbf{L})\big)_j, \ldots, \big(\mathbf{Z}^{(n)}_n(\boldsymbol{\mu}, \mathbf{L})\big)_j$.

The elements $g_{\mathbf{h}}$ of the generating group $\mathcal{G}^{(n)}(\mathbf{L}), \circ$ are indexed by the family $\mathcal{H}$ of k-tuples $\mathbf{h} = (h_1, \ldots, h_k)$ of monotone continuous and strictly increasing functions h_j from $\mathbb{R}$ to $\mathbb{R}$ such that $\lim_{z \to \pm\infty} h_j(z) = \pm\infty$, with $g_{\mathbf{h}} \in \mathcal{G}^{(n)}(\mathbf{L})$ defined as

$$g_{\mathbf{h}} : \mathbf{x} = (\mathbf{x}'_1, \ldots, \mathbf{x}'_n)' = \big((x_{11}, \ldots, x_{1k}), \ldots, (x_{n1}, \ldots, x_{nk})\big)' \in \mathbb{R}^{kn} \mapsto g_{\mathbf{h}}(\mathbf{x})$$

where

$$g_{\mathbf{h}}(\mathbf{x}) = \Big(\mathbf{L}\big(h_1((\mathbf{L}^{-1}\mathbf{x}_1)_1), \ldots, h_k((\mathbf{L}^{-1}\mathbf{x}_1)_k)\big)', \ldots, \mathbf{L}\big(h_1((\mathbf{L}^{-1}\mathbf{x}_n)_1), \ldots, h_k((\mathbf{L}^{-1}\mathbf{x}_n)_k)\big)'\Big)'.$$

That is, $\mathcal{G}^{(n)}(\mathbf{L}), \circ$ is a transformation-retransformation form of the group of continuous marginal order-preserving transformations acting componentwise on the $\mathbf{L}^{-1}\mathbf{X}^{(n)}_i$'s. Standard results on ranks entail that this group is generating $\mathcal{P}^{(n)}_{\mathbf{L}}$ and has maximal invariant $\mathbf{R}^{(n)}(\mathbf{L})$.

A similar situation holds when the parameter of interest is $(\boldsymbol{\mu}, \mathbf{L})$; similar ideas then lead to considering a smaller group $\mathcal{G}^{(n)}_0(\mathbf{L})$, with maximal invariant the componentwise signs and ranks extending the methods proposed in Hallin et al. (2006 and 2008). This latter approach

is not needed here, where we focus on R-estimation of $\mathbf{L}$, but it is considered in Hallin and Mehta (2013), who study testing problems for location and regression.

The approach by Ilmonen and Paindaveine (2011) is quite parallel. However, although addressing the problem of estimating the mixing matrix $\mathbf{\Lambda}$, so that $\boldsymbol{\mu}$ is a nuisance, these authors do not consider the group $\mathcal{G}^{(n)}(\mathbf{L})$, nor the group $\mathcal{G}_0^{(n)}(\mathbf{L})$. They rather make the additional assumption that the k component densities f_j all are symmetric with respect to the origin. Under that assumption, they are using yet another group, which is the subgroup $\mathcal{G}_+^{(n)}(\mathbf{L})$ of $\mathcal{G}^{(n)}(\mathbf{L})$ corresponding to those $\mathbf{h} \in \mathcal{H}$ such that $h_j(-z) = -h_j(z)$ for all $j = 1, \ldots, k$ and $z \in \mathbb{R}$. The resulting maximal invariant is a vector of componentwise *signed ranks*, that is, the vector of componentwise residual signs, along with the vector $\mathbf{R}_+^{(n)}(\boldsymbol{\mu}, \mathbf{L}) = (\mathbf{R}_{+1}^{(n)\prime}(\boldsymbol{\mu}, \mathbf{L}), \ldots, \mathbf{R}_{+n}^{(n)\prime}(\boldsymbol{\mu}, \mathbf{L}))'$, where $\mathbf{R}_{+i}^{(n)}(\boldsymbol{\mu}, \mathbf{L}) = (R_{+i1}^{(n)}(\boldsymbol{\mu}, \mathbf{L}), \ldots, R_{+ik}^{(n)}(\boldsymbol{\mu}, \mathbf{L}))'$, with $R_{+ij}^{(n)}(\boldsymbol{\mu}, \mathbf{L})$ the rank of $\left|\left(\mathbf{Z}_i^{(n)}(\boldsymbol{\mu}, \mathbf{L})\right)_j\right|$ among $\left|\left(\mathbf{Z}_1^{(n)}(\boldsymbol{\mu}, \mathbf{L})\right)_j\right|, \ldots, \left|\left(\mathbf{Z}_n^{(n)}(\boldsymbol{\mu}, \mathbf{L})\right)_j\right|$. As a result, their estimators lose root-n consistency as soon as one of the underlying f_j's fails to be symmetric with respect to zero—an assumption that hardly can be checked for.

2.2 Uniform local asymptotic normality (ULAN)

Establishing ULAN requires regularity conditions on f. The following conditions are sufficient for $f = (f_1, \ldots, f_k)$ to belong to $\mathcal{F}_{\text{ULAN}}$.

(A3) The component densities f_j, $j = 1, \ldots, k$, are *absolutely continuous*, that is, there exist k real-valued functions $\dot{f}_j$ such that, for any $a < b$, $f_j(b) - f_j(a) = \int_a^b \dot{f}_j(z)\mathrm{d}z$.

Letting $\boldsymbol{\varphi}_f(\mathbf{z}) := (\varphi_{f_1}(z_1), \ldots, \varphi_{f_k}(z_k))'$, $\mathbf{z} = (z_1, \ldots, z_k)' \in \mathbb{R}^k$, with $\varphi_{f_j} := -f_j'/f_j$, assume moreover that

(A4) all component densities f_j admit finite second-order moments, finite information for location, and finite information for scale; i.e. for $j = 1, \ldots, k$, $s_{f_j}^2 := \displaystyle\int_{-\infty}^{\infty} z^2 f_j(z)\mathrm{d}z$, $\mathcal{I}_{f_j} := \displaystyle\int_{-\infty}^{\infty} \varphi_{f_j}^2(z) f_j(z)\mathrm{d}z$, and $\mathcal{J}_{f_j} := \displaystyle\int_{-\infty}^{\infty} z^2 \varphi_{f_j}^2(z) f_j(z)\mathrm{d}z$ are finite.

For such f, it follows from the Cauchy-Schwarz inequality that $\alpha_{f_j} := \int_{-\infty}^{\infty} z f_j(z)\mathrm{d}z$ and $\kappa_{f_j} := \int_{-\infty}^{\infty} \varphi_{f_j}^2(z) z f_j(z)\mathrm{d}z$, $j = 1, \ldots, k$, also are finite. Consequently, the quantities $\gamma_{pq}(f) := \mathcal{I}_{f_p} s_{f_q}^2$, $\varsigma_{pq}(f) := \alpha_{f_p}\kappa_{f_q}$, and $\varrho_{jpq}(f) := \mathcal{I}_{f_j}\alpha_{f_p}\alpha_{f_q}$, are bounded for every $j, p, q \in \{1, \ldots, k\}$. The information matrix for the ULAN result, in Proposition 2.1 below, depends on these quantities through

$$\begin{aligned}\mathbf{G}_f \;:=\; & \sum_{j=1}^{k} \big(\mathcal{J}_{f_j} - 1\big)\big(\mathbf{e}_j\mathbf{e}_j' \otimes \mathbf{e}_j\mathbf{e}_j'\big) + \sum_{\substack{p,q=1\\ p\neq q}}^{k} \big\{\gamma_{qp}(f)\big(\mathbf{e}_p\mathbf{e}_p' \otimes \mathbf{e}_q\mathbf{e}_q'\big) + \big(\mathbf{e}_p\mathbf{e}_q' \otimes \mathbf{e}_q\mathbf{e}_p'\big)\big\} \\ & + \sum_{\substack{r,s=1\\ r\neq s}}^{k} \mathbf{e}_p\mathbf{e}_q' \otimes \big(\varsigma_{pq}(f)\mathbf{e}_q\mathbf{e}_q' + \varsigma_{qp}(f)\mathbf{e}_p\mathbf{e}_p'\big) + \sum_{\substack{j,p,q=1\\ j\neq p, j\neq q, p\neq q}}^{k} \varrho_{jpq}(f)\,\big(\mathbf{e}_p\mathbf{e}_q' \otimes \mathbf{e}_j\mathbf{e}_j'\big), \quad (2.5)\end{aligned}$$

where $\mathbf{e}_j$ is the jth canonical basis vector of $\mathbb{R}^k$ and $\otimes$ denotes the Kronecker product.

Writing $\mathbf{I}_k$ for the $k\times k$ identity matrix, define $\mathbf{C} := \sum_{p=1}^{k}\sum_{q=1}^{k-1} \mathbf{e}_p\mathbf{e}_p' \otimes \mathbf{u}_q\mathbf{e}_{q+\delta_{q\geq p}}'$, where $\mathbf{u}_q$ is the qth canonical basis vector of $\mathbb{R}^{k-1}$ and $\mathbf{e}_{q+\delta_{q\geq p}} := \delta_{q\geq p}\mathbf{e}_{q+1} + (1-\delta_{q\geq p})\mathbf{e}_q$, with $\delta_{q\geq p}$ the indicator for $q \geq p$. Then, let $\mathrm{odiag}(\mathbf{M})$ replace the diagonal entries of a matrix $\mathbf{M}$ with zeros. Finally, for any $\mathbf{m} \in \mathbb{R}^{k(k-1)}$, define $\mathrm{matd}^\circ(\mathbf{m})$ as the unique $k \times k$ matrix with a diagonal of zeroes such that $\mathrm{vecd}^\circ(\mathrm{matd}^\circ(\mathbf{m})) = \mathbf{m}$.

Proposition 2.1. *Let $f \in \mathcal{F}_0$ satisfy (A3) and (A4). Then, $f \in \mathcal{F}_{ULAN}$, and, for any fixed $\boldsymbol{\mu} \in \mathbb{R}^k$, the sequence of subfamilies $\mathcal{P}_{\boldsymbol{\mu};f}^{(n)}$, with parameter $\mathbf{L} \in \mathcal{M}_k^1$, is ULAN with central sequence*

$$\boldsymbol{\Delta}_{\mathbf{L};\,\boldsymbol{\mu},f}^{(n)} = \mathbf{C}\big(\mathbf{I}_k \otimes \mathbf{L}^{-1}\big)'\mathrm{vec}\big[\mathbf{T}_{\mathbf{L};\,\boldsymbol{\mu},f}^{(n)}\big], \quad \textit{where} \quad \mathbf{T}_{\mathbf{L};\,\boldsymbol{\mu},f}^{(n)} := n^{-\frac{1}{2}}\sum_{i=1}^{n}\big(\boldsymbol{\varphi}_f\big(\mathbf{Z}_i^{(n)}\big)\mathbf{Z}_i^{(n)\prime} - \mathbf{I}_k\big) \quad (2.6)$$

where $\mathbf{Z}_i^{(n)} := \mathbf{Z}_i^{(n)}\big(\boldsymbol{\mu}, \mathbf{L}\big)$ is defined in (2.4), and full-rank information matrix

$$\boldsymbol{\Gamma}_{\mathbf{L};f} := \mathbf{C}\big(\mathbf{I}_k \otimes \mathbf{L}^{-1}\big)'\mathbf{G}_f\big(\mathbf{I}_k \otimes \mathbf{L}^{-1}\big)\mathbf{C}', \quad (2.7)$$

with $\mathbf{G}_f$ *defined in (2.5). Specifically, for any sequence* $\mathbf{L}^{(n)} = \mathbf{L} + O(n^{-\frac{1}{2}}) \in \mathcal{M}_k^1$ *and any bounded sequence* $\boldsymbol{\tau}^{(n)} \in \mathbb{R}^{k(k-1)}$,

$$\log \frac{\mathrm{dP}^{(n)}_{\boldsymbol{\mu},\mathbf{L}^{(n)}+n^{-\frac{1}{2}}\mathrm{matd}^\circ(\boldsymbol{\tau}^{(n)});f}}{\mathrm{dP}^{(n)}_{\boldsymbol{\mu},\mathbf{L}^{(n)};f}} = \boldsymbol{\tau}^{(n)\prime}\boldsymbol{\Delta}^{(n)}_{\mathbf{L}^{(n)};\,\boldsymbol{\mu},f} - \frac{1}{2}\boldsymbol{\tau}^{(n)\prime}\boldsymbol{\Gamma}_{\mathbf{L};f}\boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1) \tag{2.8}$$

and $\boldsymbol{\Delta}^{(n)}_{\mathbf{L}^{(n)};\,\boldsymbol{\mu},f} \xrightarrow{\mathcal{L}} \mathcal{N}_{k(k-1)}\big(\mathbf{0}, \boldsymbol{\Gamma}_{\mathbf{L};f}\big)$, *as* $n \to \infty$ *under* $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L}^{(n)};f}$.

This ULAN property extends that established by Oja et al. (2009) under the additional assumption that each component density f_j is symmetric. Symmetry for every f_j implies that the quantities α_{f_j} and κ_{f_j}, hence also the quantities ς_{jp} and ϱ_{jpq}, all take value zero for $j, p, q \in \{1, \ldots, k\}$; therefore, dropping this assumption of symmetry affects the information matrix (2.7) through $\mathbf{G}_f$ in (2.5), which explains why our $\boldsymbol{\Gamma}_{\mathbf{L};f}$ differs from theirs.

2.3 Rank-based versions of central sequences

The ULAN result from Proposition 2.1 allows the construction of parametrically efficient inference procedures on the mixing matrix $\mathbf{L} \in \mathcal{M}_k^1$ at any given f and $\boldsymbol{\mu}$. In practice, these two nuisance parameters are not known. In general, misspecifying either or both of them leads to invalid inference—tests that fail to reach the nominal asymptotic level and estimators that do not achieve root-n consistency. Therefore, the semiparametric approach under which both f and $\boldsymbol{\mu}$ are unspecified is the most sensible one.

The standard semiparametric approach to the problem is the tangent space project method described in the monograph by by Bickel et al. (1993). That approach has been taken by Chen and Bickel (2006) and involves estimating the k component density scores—consequently, its effectiveness is mitigated in the absence of large sample sizes or in the presence of outliers.

As in Ilmonen and Paindaveine (2011), we consider, instead, the result of Hallin and Werker (2003) showing that, under very general conditions, the parametric central sequence

conditioned on the maximal invariant mentioned in (A2) is a version (central sequences are always defined up to $o_{\mathrm{P}}(1)$ quantities) of the semiparametrically efficient central sequence based on the tangent space projection. Our maximal invariants, however, are not the same.

Let $\mathbf{F} : \mathbb{R}^k \longrightarrow [0,1]^k$ and $\mathbf{J}_f : [0,1]^k \longrightarrow \mathbb{R}^k$ be defined so that, for $\mathbf{z} = (z_1, \ldots, z_k)' \in \mathbb{R}^k$, $\mathbf{F}(\mathbf{z}) := (F_1(z_1), \ldots, F_k(z_k))'$, with $F_j(z_j) := \int_{-\infty}^{z_j} f_j(z)\mathrm{d}z$ for $j = 1, \ldots, k$, and, for $\mathbf{u} = (u_1, \ldots, u_k)' \in [0,1]^k$, $\mathbf{J}_f(\mathbf{u}) := \boldsymbol{\varphi}_f\left(\mathbf{F}^{-1}(\mathbf{u})\right) = \left(\varphi_{f_1}\left(F_1^{-1}(u_1)\right), \ldots, \varphi_{f_k}\left(F_k^{-1}(u_k)\right)\right)'$, with $J_{f_j}(u_j) := \varphi_{f_j}\left(F_j^{-1}(u_j)\right)$ for $j = 1, \ldots, k$. Writing $\mathbf{U}_i^{(n)} := (U_{i,1}^{(n)}, \ldots, U_{i,k}^{(n)})'$ for $\mathbf{U}_i^{(n)}(\boldsymbol{\mu}, \mathbf{L}) := \mathbf{F}\left(\mathbf{Z}_i^{(n)}(\boldsymbol{\mu}, \mathbf{L})\right)$, $i = 1, \ldots, n$, the parametric statistic $\mathbf{T}^{(n)}_{\mathbf{L};\,\boldsymbol{\mu},f}$ defined in (2.6) takes the form

$$\mathbf{T}^{(n)}_{\mathbf{L};\,\boldsymbol{\mu},f} = n^{-\frac{1}{2}} \sum_{i=1}^{n} \left(\mathbf{J}_f(\mathbf{U}_i^{(n)})\mathbf{F}^{-1\prime}(\mathbf{U}_i^{(n)}) - \mathbf{I}_k\right).$$

Assume moreover that

(A5) for all $j = 1, \ldots, k$, $z \mapsto \varphi_{f_j}(z)$ is the difference of two monotone increasing functions.

Assumption (A5) will be required whenever rank-based statistics with scores $\varphi_{f_j} \circ F_j^{-1}$ are considered. Conditioning $\boldsymbol{\Delta}^{(n)}_{\mathbf{L};\,\boldsymbol{\mu},f}$ on the sigma-field $\mathcal{B}(\mathbf{L})$ generated by the marginal ranks of the $\mathbf{Z}_i^{(n)}(\boldsymbol{\mu}, \mathbf{L})$'s yields

$$\begin{aligned}
\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{L};f:\mathrm{ex}} &:= \mathrm{E}\left[\boldsymbol{\Delta}^{(n)}_{\mathbf{L};\,\boldsymbol{\mu},f} | \mathcal{B}(\mathbf{L})\right] \\
&= \mathbf{C}\left(\mathbf{I}_k \otimes \mathbf{L}^{-1}\right)' \mathrm{vec}\left[\underset{\sim}{\mathbf{T}}^{(n)}_{\mathbf{L};f:\mathrm{ex}}\right] \quad \text{where} \quad \underset{\sim}{\mathbf{T}}^{(n)}_{\mathbf{L};f:\mathrm{ex}} := \mathrm{E}\left[\mathbf{T}^{(n)}_{\mathbf{L};\,\boldsymbol{\mu},f} \middle| \mathcal{B}(\mathbf{L})\right]; \qquad (2.9)
\end{aligned}$$

clearly, $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{L};f:\mathrm{ex}}$ does not depend on $\boldsymbol{\mu}$. Computing this conditional expectation requires evaluating, for each $j \in \{1, \ldots k\}$ and $r \in \{1, \ldots, n\}$,

$$\mathrm{E}\left[J_{f_j}(U_{i,j}^{(n)}) F_j^{-1}(U_{i,j}^{(n)}) | R_{i,j}^{(n)}(\mathbf{L}) = r\right] = \mathrm{E}\left[J_{f_j}(U_{(r)}^{(n)}) F_j^{-1}(U_{(r)}^{(n)})\right] \qquad (2.10)$$

and, for each $j' \neq j'' \in \{1, \ldots k\}$ and $r, s \in \{1, \ldots, n\}$,

$$\mathrm{E}\left[J_{f_{j'}}(U_{i,j'}^{(n)})F_{j''}^{-1}(U_{i,j''}^{(n)})|R_{i,j'}^{(n)}(\mathbf{L})=r,\ R_{i,j''}^{(n)}(\mathbf{L})=s\right]=\mathrm{E}\left[J_{f_{j'}}(U_{(r)}^{(n)})\right]\mathrm{E}\left[F_{j''}^{-1}(U_{(s)}^{(n)})\right],\quad(2.11)$$

where $U_{(r)}^{(n)}$ and $U_{(s)}^{(n)}$ respectively denote, in a sample $U_1,\ldots,U_n$ of i.i.d. random variables uniform over $(0,1)$, the rth and sth order statistics. As a function of r and s, such quantities are called *exact* scores; they depend on n, and computing them via numerical integration is somewhat tedious.

The so-called *approximate scores*, in general, are preferable: denoting by

$$\underset{\sim}{\mathbf{R}}_i^{(n)}(\mathbf{L}):=\left(\underset{\sim}{R}_{i,1}^{(n)}(\mathbf{L}),\ldots,\underset{\sim}{R}_{i,k}^{(n)}(\mathbf{L})\right)':=\left(\frac{R_{i,1}^{(n)}(\mathbf{L})}{n+1},\ldots,\frac{R_{i,k}^{(n)}(\mathbf{L})}{n+1}\right)'$$

the (marginal) *normalized ranks*, the approximate scores corresponding to (2.10) and (2.11) are

$$\begin{aligned}&J_{f_j}\big(\underset{\sim}{R}_{i,j}^{(n)}(\mathbf{L})\big)F_j^{-1}\big(\underset{\sim}{R}_{i,j}^{(n)}(\mathbf{L})\big)-\frac{1}{n}\sum_{i=1}^n J_{f_j}\Big(\frac{i}{n+1}\Big)F_j^{-1}\Big(\frac{i}{n+1}\Big)\quad\text{and}\\ &J_{f_{j'}}\big(\underset{\sim}{R}_{i,j'}^{(n)}(\mathbf{L})\big)F_{j''}^{-1}\big(\underset{\sim}{R}_{i,j''}^{(n)}(\mathbf{L})\big)-\frac{1}{n}\sum_{i=1}^n J_{f_{j'}}\Big(\frac{i}{n+1}\Big)\frac{1}{n}\sum_{i=1}^n F_{j''}^{-1}\Big(\frac{i}{n+1}\Big),\end{aligned}\qquad(2.12)$$

respectively. Letting $\mathbf{1}_k\in\mathbb{R}^k$ be the k-dimensional vector of ones, the approximate-score version of the central sequence is thus

$$\underset{\sim}{\mathbf{\Delta}}_{\mathbf{L};f}^{(n)}:=\mathbf{C}\left(\mathbf{I}_k\otimes\mathbf{L}^{-1}\right)'\mathrm{vec}\Big[\underset{\sim}{\mathbf{T}}_{\mathbf{L};f}^{(n)}\Big],\qquad\text{where}\qquad(2.13)$$

$$\underset{\sim}{\mathbf{T}}_{\mathbf{L};f}^{(n)}:=\mathrm{odiag}\Big[n^{-\frac{1}{2}}\sum_{i=1}^n\Big(\mathbf{J}_f\big(\underset{\sim}{\mathbf{R}}_i^{(n)}(\mathbf{L})\big)\mathbf{F}^{-1\prime}\big(\underset{\sim}{\mathbf{R}}_i^{(n)}(\mathbf{L})\big)-\overline{\mathbf{J}}_f^{(n)}\overline{\mathbf{F}^{-1}}^{(n)\prime}\Big)\Big]\qquad(2.14)$$

with $\overline{\mathbf{J}}_f^{(n)}:=\frac{1}{n}\sum_{i=1}^n\mathbf{J}_f\Big(\frac{i}{n+1}\mathbf{1}_k\Big)$ and $\overline{\mathbf{F}^{-1}}^{(n)}:=\frac{1}{n}\sum_{i=1}^n\mathbf{F}^{-1}\Big(\frac{i}{n+1}\mathbf{1}_k\Big)$.

The following proposition, by establishing the asymptotic equivalence between the exact- and approximate-score forms (2.9) and (2.13), shows that (2.13) indeed is a version of the semiparametrically efficient central sequence for the problem.

Proposition 2.2. *Fix $\boldsymbol{\mu}\in\mathbb{R}^k$, $\mathbf{L}\in\mathcal{M}_k^1$, and $f\in\mathcal{F}_{ULAN}$ satisfying (A5). Then, under $\mathrm{P}_{\boldsymbol{\mu},\mathbf{L};f}^{(n)}$,*

$$(i)\ \underset{\sim}{\mathbf{\Delta}}_{\mathbf{L};f}^{(n)}=\underset{\sim}{\mathbf{\Delta}}_{\mathbf{L};f:\mathrm{ex}}^{(n)}+o_{L^2}(1)\qquad\text{and}\qquad(ii)\ \underset{\sim}{\mathbf{\Delta}}_{\mathbf{L};f}^{(n)}=\mathbf{\Delta}_{\mathbf{L},\boldsymbol{\mu};f}^{(n)*}+o_{L^2}(1),$$

as $n\to\infty$, where $\mathbf{\Delta}_{\mathbf{L},\boldsymbol{\mu};f}^{(n)*}$ is a semiparametrically efficient (at $\mathbf{L}$, $\boldsymbol{\mu}$, and f) central sequence.

Consequently, $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{L};f}$ can be used to construct semiparametrically efficient (at f, irrespective of $\boldsymbol{\mu}$) estimation procedures for $\mathbf{L}$. Contrary to those based on $\boldsymbol{\Delta}^{(n)*}_{\mathbf{L},\boldsymbol{\mu};f}$, the R-estimators derived from $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{L};f}$ remain root-n consistent, though, under most component densities $g \in \mathcal{F}_{\text{ULAN}}$, $g \neq f$. And, unlike those proposed by Ilmonen and Paindaveine (2011), they do not require f nor g to be symmetric.

The asymptotic representation for the rank-based central sequence $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{L};f}$ under $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L};g}$ where $g \in \mathcal{F}_0$ is not necessarily equal to $f \in \mathcal{F}_{\text{ULAN}}$ is described in the next proposition. If, additionally, $g \in \mathcal{F}_{\text{ULAN}}$, the asymptotic distribution for $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{L};f}$ can be made explicit. For every $p \neq q \in \{1,\ldots,k\}$, let

$$\begin{aligned} \gamma^*_{pq}(f,g) &:= \int_0^1 \varphi_{f_p}\big(F_p^{-1}(u)\big)\varphi_{g_p}\big(G_p^{-1}(u)\big)\mathrm{d}u\Big(\int_0^1 F_q^{-1}(u)G_q^{-1}(u)\mathrm{d}u - \alpha_{f_q}\alpha_{g_q}\Big) \\ \text{and} \quad \rho^*_{pq}(f,g) &:= \int_0^1 F_p^{-1}(u)\varphi_{f_p}\big(G_p^{-1}(u)\big)\mathrm{d}u \int_0^1 \varphi_{f_q}\big(F_q^{-1}(u)\big)G_q^{-1}(u)\mathrm{d}u. \end{aligned}$$

The quantities $\gamma^*_{pq}(f,g)$ and $\rho^*_{pq}(f,g)$ are referred to as *cross-information* quantities; note that $\gamma^*_{pq}(f,f) = \gamma_{pq}(f) - \varrho_{pqq}(f)$ and $\rho^*_{pq}(f,f) = 1$. Then, define

$$\boldsymbol{\Gamma}^*_{\mathbf{L};f,g} := \mathbf{C}\left(\mathbf{I}_k \otimes \mathbf{L}^{-1}\right)' \tilde{\mathbf{G}}_{f,g}\left(\mathbf{I}_k \otimes \mathbf{L}^{-1}\right)\mathbf{C}' \tag{2.15}$$

where $\tilde{\mathbf{G}}_{f,g} := \sum_{p\neq q=1}^k \gamma^*_{sr}(f,g)\big(\mathbf{e}_p\mathbf{e}_p' \otimes \mathbf{e}_q\mathbf{e}_q'\big) + \rho^*_{pq}(f,g)\big(\mathbf{e}_p\mathbf{e}_q' \otimes \mathbf{e}_q\mathbf{e}_p'\big)$, and write $\boldsymbol{\Gamma}^*_{\mathbf{L},f}$ for $\boldsymbol{\Gamma}^*_{\mathbf{L},f,f}$. Remark that $\boldsymbol{\Gamma}^*_{\mathbf{L},f,g}$ depends on g only through $\gamma^*_{pq}(f,g)$ and $\rho^*_{pq}(f,g)$.

Proposition 2.3. *Fix $f \in \mathcal{F}_{ULAN}$, $\boldsymbol{\mu} \in \mathbb{R}^k$, and $\mathbf{L} \in \mathcal{M}^1_k$; with $\mathbf{Z}^{(n)}_i := \mathbf{Z}^{(n)}_i\big(\boldsymbol{\mu},\mathbf{L}\big)$ defined in (2.4), let $\widetilde{\mathbf{J}}^{(n)}_f := \frac{1}{n}\sum_{i=1}^n \mathbf{J}_f\big(\mathbf{G}\big(\mathbf{Z}^{(n)}_i\big)\big)$ and $\widetilde{\mathbf{F}}^{-1(n)} := \frac{1}{n}\sum_{i=1}^n \mathbf{F}^{-1}\big(\mathbf{G}\big(\mathbf{Z}^{(n)}_i\big)\big)$. Then,*

(i) If $g \in \mathcal{F}_0$, $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{L};f} = \boldsymbol{\Delta}^{\diamond(n)}_{\mathbf{L},\boldsymbol{\mu};f,g} + o_{L^2}(1)$ as $n \to \infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L},g}$, where

$$\boldsymbol{\Delta}^{\diamond(n)}_{\mathbf{L},\boldsymbol{\mu};f,g} := \mathbf{C}\left(\mathbf{I}_k \otimes \mathbf{L}^{-1}\right)' \mathrm{vec}\Big[\mathbf{T}^{\diamond(n)}_{\mathbf{L},\boldsymbol{\mu};f,g}\Big],$$

and

$$\mathbf{T}^{\diamond(n)}_{\mathbf{L},\boldsymbol{\mu};f,g} := \mathrm{odiag}\big[n^{-\frac{1}{2}}\sum_{i=1}^{n}\big(\mathbf{J}_f(\mathbf{G}(\mathbf{Z}^{(n)}_i))\mathbf{F}^{-1\prime}(\mathbf{G}(\mathbf{Z}^{(n)}_i)) - \widetilde{\mathbf{J}}^{(n)}_f\widetilde{\mathbf{F}}^{-1(n)\prime}\big)\big]. \tag{2.16}$$

(ii) Suppose furthermore that $g \in \mathcal{F}_{ULAN}$*, and fix* $\boldsymbol{\tau} \in \mathbb{R}^{k(k-1)}$ *so that* $\mathbf{L}+n^{-\frac{1}{2}}\mathrm{matd}^\circ(\boldsymbol{\tau}) \in \mathcal{M}^1_k$*. Then,* $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{L};f} \xrightarrow{\mathcal{L}} \mathcal{N}_{k(k-1)}\big(\boldsymbol{\Gamma}^*_{\mathbf{L};f,g}\boldsymbol{\tau}, \boldsymbol{\Gamma}^*_{\mathbf{L};\mathbf{f}}\big)$ *as* $n\to\infty$*, under* $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L}+n^{-\frac{1}{2}}\mathrm{matd}^\circ(\boldsymbol{\tau});g}$ *with* $\boldsymbol{\Gamma}^*_{\mathbf{L};f,g}$ *defined in (2.15). If* $\boldsymbol{\tau}=\mathbf{0}_{k(k-1)}$*,* $g\in\mathcal{F}_0$ *is sufficient for this convergence to hold.*

(iii) If, again, $g\in\mathcal{F}_{ULAN}$ *and* $\boldsymbol{\tau}\in\mathbb{R}^{k(k-1)}$ *is as defined in (ii), then, as* $n\to\infty$*, under* $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L};g}$*,*

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{L}+n^{-\frac{1}{2}}\mathrm{matd}^\circ(\boldsymbol{\tau});f} - \ \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{L};f} = -\boldsymbol{\Gamma}^*_{\mathbf{L};f,g}\boldsymbol{\tau} + o_{\mathrm{P}}(1). \tag{2.17}$$

In Section 3, our R-estimation procedures require evaluating the f-score rank-based central sequence, for $f\in\mathcal{F}_{\mathrm{ULAN}}$, at a preliminary root-$n$ consistent estimator $\tilde{\mathbf{L}}^{(n)}$ of $\mathbf{L}$. The asymptotic impact of substituting $\tilde{\mathbf{L}}^{(n)}$ for $\mathbf{L}$ does not directly follow from Proposition 2.3(iii) because the perturbation $\boldsymbol{\tau}$ in (2.17) is a deterministic quantity. Lemma 4.4 in Kreiss (1987) provides sufficient conditions for Proposition 2.3(iii) to hold when replacing $\boldsymbol{\tau}$ with a sequence of random vectors, $\tilde{\boldsymbol{\tau}}^{(n)}$, $n\in\mathbb{N}$. More precisely, if

(C1a) $\tilde{\boldsymbol{\tau}}^{(n)} = O_{\mathrm{P}}(1)$, as $n\to\infty$, and

(C1b) there exists an integer $N<\infty$ so that, for all $n\geq N$, $\tilde{\boldsymbol{\tau}}^{(n)}$ can take, at most, a finite number of values within any bounded ball centered at the origin in $\mathbb{R}^{k(k-1)}$,

hold, then (2.17) is still valid with $\boldsymbol{\tau}$ replaced by $\tilde{\boldsymbol{\tau}}^{(n)}$.

Let $\tilde{\mathbf{L}}^{(n)}\in\mathcal{M}^1_k$ be an estimator for $\mathbf{L}$. We say that it is *root-n consistent* under $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L};g}$ and *locally asymptotically discrete* if $n^{\frac{1}{2}}\mathrm{vecd}^\circ\big(\tilde{\mathbf{L}}^{(n)}-\mathbf{L}\big)$ satisfies (C1a) under $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L};g}$ and (C1b). Proposition 2.3(iii) and Lemma 4.4 from Kreiss (1987) then yield the following corollary.

Corollary 2.1. *Fix* $\boldsymbol{\mu}\in\mathbb{R}^k$*,* $\mathbf{L}\in\mathcal{M}^1_k$ *and* $f,g\in\mathcal{F}_{ULAN}$*. Suppose that* $\tilde{\mathbf{L}}^{(n)}$ *is root-n consistent*

under $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L};g}$ *and locally asymptotically discrete. Then, under* $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L};g}$*, as* $n \to \infty$*,*

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\tilde{\mathbf{L}}^{(n)},f} - \ \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{L},f} = -\boldsymbol{\Gamma}^{*}_{\mathbf{L},f,g}\mathrm{vecd}^{\circ}\big(\tilde{\mathbf{L}}^{(n)} - \mathbf{L}\big) + o_{\mathrm{P}}(1). \tag{2.18}$$

The asymptotic discreteness requirement for the preliminary estimator is not overly restrictive. Any root-n consistent sequence $\tilde{\mathbf{L}}^{(n)} := (\tilde{L}^{(n)}_{rs}) \in \mathcal{M}^1_k$ indeed can be discretized as $\tilde{\mathbf{L}}^{(n)}_{\#} := (\tilde{L}^{(n)}_{rs;\#})$, with $\tilde{L}^{(n)}_{rs;\#} := \big(cn^{\frac{1}{2}}\big)^{-1}\mathrm{sign}\big(\tilde{L}^{(n)}_{rs}\big)\Big\lceil cn^{\frac{1}{2}}\big|\tilde{L}^{(n)}_{rs}\big|\Big\rceil$, for $r \neq s \in \{1,\ldots,k\}$, where $c > 0$ is an arbitrary constant and $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. The root-n consistency properties of $\tilde{\mathbf{L}}^{(n)}$ carry over to $\tilde{\mathbf{L}}^{(n)}_{\#}$ which by construction is locally asymptotically discrete and, because $\mathcal{M}^1_k$ is a compact subset of $\mathbb{R}^{k(k-1)}$, still takes values in $\mathcal{M}^1_k$.

3 R-estimation of the mixing matrix

Assume that a rank test rejects $\mathrm{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against the alternative $\mathrm{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ for large values of some test statistic $Q_{\boldsymbol{\theta}_0}\big(\mathbf{R}^{(n)}\big(\boldsymbol{\theta}_0\big)\big)$ measurable with respect to the ranks $\mathbf{R}^{(n)}\big(\boldsymbol{\theta}_0\big)$ of residuals $\mathbf{Z}^{(n)}(\boldsymbol{\theta}_0) := \big(\mathbf{Z}^{(n)}_1(\boldsymbol{\theta}_0),\ldots,\mathbf{Z}^{(n)}_n(\boldsymbol{\theta}_0)\big)'$, which are i.i.d. if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. The original R-estimator for $\boldsymbol{\theta} \in \Theta$, as proposed by Hodges and Lehmann (1963), is defined as $\hat{\boldsymbol{\theta}}^{(n)}_{\mathrm{HL}} := \arg\min_{\boldsymbol{\theta}\in\Theta} Q^{(n)}_{\boldsymbol{\theta}}\left(\mathbf{R}^{(n)}\left(\boldsymbol{\theta}\right)\right)$.

Even for simple problems such as location, regression, etc. involving a low-dimensional parameter $\boldsymbol{\theta}$, minimizing $Q^{(n)}_{\boldsymbol{\theta}}\big(\mathbf{R}^{(n)}\big(\boldsymbol{\theta}\big)\big)$ is wrought with difficulty—as a function of $\boldsymbol{\theta}$, it is piecewise constant, discontinuous, and generally non-convex. In the present case of a $k(k-1)$-dimensional parameter space $\mathcal{M}^1_k$, solving this problem typically would require an infeasible grid-search in relatively high dimension.

As an alternative, we consider the one-step R-estimators described in Hallin, Oja and Paindaveine (2006) and Hallin and Paindaveine (2013) that possess advantageous features such as expedient computation, straightforward asymptotic properties, and provide a con-

sistent estimator for the asymptotic covariance matrix as a by-product. Those one-step estimators are computed from a preliminary root-n consistent estimator $\tilde{\mathbf{L}}^{(n)}$ and the resulting value $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{\tilde{\mathbf{L}}^{(n)};f}$ of the rank-based central sequence associated with some reference density $f \in \mathcal{F}_{\text{ULAN}}$ satisfying (A5).

3.1 One-step R-estimation

For fixed $f \in \mathcal{F}_{\text{ULAN}}$, assume that

(C1) there exists a sequence of estimators $\tilde{\mathbf{L}}^{(n)} \in \mathcal{M}^1_k$ of the parameter $\mathbf{L} \in \mathcal{M}^1_k$ that are both root-n consistent and locally asymptotically discrete, under $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L};g}$ for any $\boldsymbol{\mu} \in \mathbb{R}^k$, $\mathbf{L} \in \mathcal{M}^1_k$, and $g \in \mathcal{F}_{\text{ULAN}}$, and, furthermore,

(C2) for all $p \neq q \in \{1, \ldots, k\}$, there exist consistent (under $\mathcal{P}^{(n)}_g$ for every $g \in \mathcal{F}_{\text{ULAN}}$) and locally asymptotically discrete sequences $\hat{\gamma}^*_{pq}(f)$ and $\hat{\rho}^*_{pq}(f)$ of estimators for the cross-information quantities $\gamma^*_{pq}(f,g)$ and $\rho^*_{pq}(f,g)$.

For any $f \in \mathcal{F}_{\text{ULAN}}$, the one-step R-estimator for $\mathbf{L} \in \mathcal{M}^1_k$ based on f-scores is the $k \times k$ matrix $\underset{\sim}{\mathbf{L}}^{(n)}_f \in \mathcal{M}^1_k$ defined by

$$\text{vecd}^\circ\big(\underset{\sim}{\mathbf{L}}^{(n)}_f\big) = \text{vecd}^\circ\big(\tilde{\mathbf{L}}^{(n)}\big) + n^{-\frac{1}{2}}\big(\hat{\mathbf{\Gamma}}^*_{\tilde{\mathbf{L}}^{(n)};f}\big)^{-1}\ \underset{\sim}{\mathbf{\Delta}}_{\tilde{\mathbf{L}}^{(n)};f}, \tag{3.19}$$

where $\hat{\mathbf{\Gamma}}^*_{\tilde{\mathbf{L}}^{(n)};f}$ is a consistent estimate of $\mathbf{\Gamma}^*_{\mathbf{L};f,g}$. This estimator is constructed by plugging $\hat{\gamma}^*_{pq}(f)$ and $\hat{\rho}^*_{pq}(f)$ into (2.15). Under Assumptions (C1) and (C2), $\hat{\mathbf{\Gamma}}^*_{\tilde{\mathbf{L}}^{(n)};f}$ is a consistent estimate of $\mathbf{\Gamma}^*_{\mathbf{L};f,g}$. The procedure for obtaining each estimate $\hat{\gamma}^*_{pq}(f)$ and $\hat{\rho}^*_{pq}(f)$ satisfying (C2) is discussed in Section 3.2. The next proposition establishes the asymptotic distribution of $\underset{\sim}{\mathbf{L}}^{(n)}_f$; its proof parallels that of Theorem 5.1 in Ilmonen and Paindaveine (2011).

Proposition 3.1. *Fix a reference density $f \in \mathcal{F}_{ULAN}$. Then,*

(i) for any $\boldsymbol{\mu} \in \mathbb{R}^k$, $\mathbf{L} \in \mathcal{M}_k^1$, *and* $g \in \mathcal{F}_{ULAN}$, *the one-step R-estimator (3.19) is such that*

$$n^{\frac{1}{2}} \mathrm{vecd}^\circ \left(\underset{\sim}{\mathbf{L}}_f^{(n)} - \mathbf{L} \right) \xrightarrow{\mathcal{L}} \mathcal{N}_{k(k-1)} \left(\mathbf{0}, \left(\boldsymbol{\Gamma}^*_{\mathbf{L};f,g} \right)^{-1} \boldsymbol{\Gamma}^*_{\mathbf{L};f} \left(\boldsymbol{\Gamma}^*_{\mathbf{L};f,g} \right)^{-1} \right) \tag{3.20}$$

as $n \to \infty$, *under* $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L},g}$, *and*

(ii) if, moreover, $f = g$, *then* $\left(\boldsymbol{\Gamma}^*_{\mathbf{L};f,g} \right)^{-1} \boldsymbol{\Gamma}^*_{\mathbf{L};f} \left(\boldsymbol{\Gamma}^*_{\mathbf{L};f,g} \right)^{-1} = \left(\boldsymbol{\Gamma}^*_{\mathbf{L};f} \right)^{-1}$, *and* $\underset{\sim}{\mathbf{L}}_f^{(n)}$ *is a semi-parametrically efficient (at* f*) estimate of* $\mathbf{L}$.

The *R*-estimator $\underset{\sim}{\mathbf{L}}_f^{(n)}$ can be written in a form that avoids inverting $\hat{\boldsymbol{\Gamma}}^*_{\tilde{\mathbf{L}}^{(n)};f}$, which can be numerically singular when estimated in practice. Define therefore the $k \times k$ matrices $\hat{\mathcal{A}}^{(n)}_{\tilde{\mathbf{L}}^{(n)},f} := (\hat{\alpha}_{pq}(f))^k_{p,q=1}$ and $\hat{\mathcal{B}}^{(n)}_{\tilde{\mathbf{L}}^{(n)},f} := (\hat{\beta}_{pq}(f))^k_{p,q=1}$ with zeroes on the diagonal and, for every $p \neq q \in \{1, \ldots, k\}$,

$$\hat{\alpha}^{(n)}_{pq}(f) := \frac{\hat{\gamma}^*_{pq}(f)}{\hat{\gamma}^*_{pq}(f)\hat{\gamma}^*_{qp}(f) - \hat{\rho}^*_{pq}(f)\hat{\rho}^*_{qp}(f)} \quad \text{and} \quad \hat{\beta}^{(n)}_{pq}(f) := \frac{-\hat{\rho}^*_{pq}(f)}{\hat{\gamma}^*_{pq}(f)\hat{\gamma}^*_{qp}(f) - \hat{\rho}^*_{pq}(f)\hat{\rho}^*_{qp}(f)}.$$

Letting $\mathbf{A} \odot \mathbf{B} = (a_{pq}b_{pq})$ denote the Hadamard product between two matrices $\mathbf{A} = (a_{pq})$ and $\mathbf{B} = (b_{pq})$ of the same size, define

$$\hat{\mathbf{N}}^{(n)}_{\tilde{\mathbf{L}}^{(n)},f} := \left(\hat{\mathcal{A}}^{(n)\prime}_{\tilde{\mathbf{L}}^{(n)},f} \odot \underset{\sim}{\mathbf{T}}^{(n)}_{\tilde{\mathbf{L}}^{(n)};f} \right) + \left(\hat{\mathcal{B}}^{(n)\prime}_{\tilde{\mathbf{L}}^{(n)},f} \odot \underset{\sim}{\mathbf{T}}^{(n)\prime}_{\tilde{\mathbf{L}}^{(n)};f} \right), \tag{3.21}$$

with $\underset{\sim}{\mathbf{T}}^{(n)}_{\mathbf{L};f}$ defined in (2.14). Theorem 5.2 in Ilmonen and Paindaveine (2011) then implies that $\underset{\sim}{\mathbf{L}}_f^{(n)}$ can be expressed as

$$\underset{\sim}{\mathbf{L}}_f^{(n)} = \tilde{\mathbf{L}}^{(n)} + n^{-\frac{1}{2}} \tilde{\mathbf{L}}^{(n)} \left[\hat{\mathbf{N}}^{(n)}_{\tilde{\mathbf{L}}^{(n)},f} - \mathrm{diag}\big(\tilde{\mathbf{L}}^{(n)} \hat{\mathbf{N}}^{(n)}_{\tilde{\mathbf{L}}^{(n)},f} \big) \right]. \tag{3.22}$$

3.2 Consistent estimation of cross-information quantities

A critical point in computing $\underset{\sim}{\mathbf{L}}_f^{(n)}$ (3.22) is the consistent estimation of the cross-information quantities in $\mathbf{\Gamma}^*_{\mathbf{L},f;g}$. To tackle this issue, we exploit the asymptotic linearity (2.17) of $\underset{\sim}{\mathbf{\Delta}}_{\tilde{\mathbf{L}}^{(n)};f}$ using a method first proposed by Hallin et al. (2006) in the context of the R-estimation of a scatter matrix in an elliptical model, and further developed by Cassart et al. (2010) and Hallin and Paindaveine (2013). In the present case, we have to consistently estimate a total of $2k(k-1)$ cross-information quantities appearing in $\mathbf{\Gamma}^*{}_{\mathbf{L};f,g}$.

Fixing $f \in \mathcal{F}_{\text{ULAN}}$, define, for $\lambda \geq 0$ and $r \neq s \in \{1,\ldots,k\}$, the mappings

$$\lambda \mapsto h^{\gamma^*_{rs}}(\lambda) := \big(\underset{\sim}{\mathbf{T}}^{(n)}_{\tilde{\mathbf{L}}^{(n)};f}\big)_{rs}\big(\underset{\sim}{\mathbf{T}}^{(n)}_{\tilde{\mathbf{L}}^{\gamma^*_{rs}}_{\lambda};f}\big)_{rs} \quad \text{and} \quad \lambda \mapsto h^{\rho^*_{rs}}(\lambda) := \big(\underset{\sim}{\mathbf{T}}^{(n)}_{\tilde{\mathbf{L}}^{(n)};f}\big)_{sr}\big(\underset{\sim}{\mathbf{T}}^{(n)}_{\tilde{\mathbf{L}}^{\rho_{rs}}_{\lambda};f}\big)_{sr} \quad (3.23)$$

(from $\mathbb{R}_+$ to $\mathbb{R}$), where

$$\begin{aligned}
\tilde{\mathbf{L}}^{\gamma^*_{rs}}_{\lambda} &:= \tilde{\mathbf{L}}^{(n)} + n^{-\frac{1}{2}}\lambda\big(\underset{\sim}{\mathbf{T}}^{(n)}_{\tilde{\mathbf{L}}^{(n)},f}\big)_{rs}\tilde{\mathbf{L}}^{(n)}\big(\mathbf{e}_r\mathbf{e}_s' - \text{diag}\big(\tilde{\mathbf{L}}^{(n)}\mathbf{e}_r\mathbf{e}_s'\big)\big) \quad \text{and} \\
\tilde{\mathbf{L}}^{\rho^*_{rs}}_{\lambda} &:= \tilde{\mathbf{L}}^{(n)} + n^{-\frac{1}{2}}\lambda\big(\underset{\sim}{\mathbf{T}}^{(n)}_{\tilde{\mathbf{L}}^{(n)},f}\big)_{sr}\tilde{\mathbf{L}}^{(n)}\big(\mathbf{e}_r\mathbf{e}_s' - \text{diag}\big(\tilde{\mathbf{L}}^{(n)}\mathbf{e}_r\mathbf{e}_s'\big)\big),
\end{aligned}$$

with $\underset{\sim}{\mathbf{T}}^{(n)}_{\mathbf{L},f}$ defined in (2.14). Assume, additionally, that

(C3) for fixed $f, g \in \mathcal{F}_{\text{ULAN}}$, $\boldsymbol{\mu} \in \mathbb{R}^k$, and $\mathbf{L} \in \mathcal{M}^1_k$, the sequence $\tilde{\mathbf{L}}^{(n)}$ of preliminary estimators (satisfying (C1)) is such that each element in $\underset{\sim}{\mathbf{T}}^{(n)}_{\tilde{\mathbf{L}}^{(n)};f}$ is bounded from below by a positive constant with probability tending to one under $\text{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L};g}$. More precisely, for all $\epsilon > 0$, there exist $\delta_\epsilon > 0$ and an integer N_ϵ such that $\text{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L},g}\left[\big(\mathbf{T}_{\tilde{\mathbf{L}}^{(n)};f}\big)_{rs} > \delta_\epsilon\right] \geq 1-\epsilon$ for all $n \geq N_\epsilon$ and $r \neq s \in \{1,\ldots,k\}$.

This assumption is satisfied by most root-n consistent estimators for the mixing matrix; see Section 4 for a discussion.

The following lemma is adapted from Hallin and Paindaveine (2013).

Lemma 3.1. *Fix $f, g \in \mathcal{F}_{ULAN}$, $\boldsymbol{\mu} \in \mathbb{R}^k$, and $\mathbf{L} \in \mathcal{M}^1_k$. Let $\tilde{\mathbf{L}}^{(n)}$ be a sequence of preliminary*

estimators for $\mathbf{L}$ *satisfying (c1) and (c3). For every* $r \neq s \in \{1, \ldots, k\}$*, the mappings* $h^{\gamma^*_{rs}}$ *and* $h^{\rho^*_{rs}}$ *defined in (3.23) satisfy, for* $\lambda > 0$*,*

$$h^{\gamma^*_{rs}}(\lambda) = (1 - \lambda\gamma^*_{rs}(f,g)) \left(\mathbf{T}_{\tilde{\mathbf{L}}^{(n)};f}\right)^2_{rs} + o_{\mathrm{P}}(1) \quad \textit{and} \quad h^{\rho^*_{rs}}(\lambda) = (1 - \lambda\rho^*_{rs}(f,g)) \left(\mathbf{T}_{\tilde{\mathbf{L}}^{(n)};f}\right)^2_{sr} + o_{\mathrm{P}}(1)$$

as $n \to \infty$*, under* $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L};g}$*. Furthermore, each mapping is almost surely positive for* $\lambda = 0$*.*

By Lemma 3.1, the mappings $h^{\gamma^*_{rs}}$ and $h^{\rho_{rs}}$ are both positive at $\lambda = 0$ and, up to $o_{\mathrm{P}}(1)$'s under $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L};g}$, are linear with a negative slope. Therefore, intuitively appealing estimators for $\gamma^*_{rs}(f,g)$ and $\rho^*_{rs}(f,g)$ would be, respectively, $\big(\hat{\gamma^*}_{rs}(f,g)\big)^{-1} := \inf_\lambda \big\{\lambda \in \mathbb{R} : h^{\gamma^*_{rs}}(\lambda) < 0\big\}$ and $\big(\hat{\rho^*}_{rs}(f,g)\big)^{-1} := \inf_\lambda \big\{\lambda \in \mathbb{R} : h^{\rho^*_{rs}}(\lambda) < 0\big\}$; estimators for $\rho_{rs}(f,g)$ would be defined in an analogous manner. However, these estimators are not asymptotically discrete. Instead, taking $\lambda_j = j/c$ for some large $c > 0$ and $j \in \mathbb{Z}$, let

$$(\hat{\gamma}^*_{rs}(f))^{-1} := \lambda^-_{\gamma^*_{rs}} + c^{-1} h^{\gamma^*_{rs}}(\lambda^-_{\gamma^*_{rs}}) / \big(h^{\gamma^*_{rs}}(\lambda^-_{\gamma^*_{rs}}) - h^{\gamma^*_{rs}}(\lambda^+_{\gamma^*_{rs}})\big), \tag{3.24}$$

with $\lambda^-_{\gamma^*_{rs}} := \max_{j\in\mathbb{Z}}\big\{\lambda_j \colon h^{\gamma^*_{rs}}(\lambda_j) > 0\big\}$ and $\lambda^+_{\gamma^*_{rs}(f)} := \min_{j\in\mathbb{Z}}\big\{\lambda_j \colon h^{\gamma^*_{rs}}(\lambda_j) < 0\big\}$. Similarly put

$$(\hat{\rho}^*_{rs}(f))^{-1} := \lambda^-_{\rho^*_{rs}} + c^{-1} h^{\rho^*_{rs}}(\lambda^-_{\rho^*_{rs}}) / \big(h^{\rho^*_{rs}}(\lambda^-_{\rho^*_{rs}}) - h^{\rho^*_{rs}}(\lambda^+_{\rho^*_{rs}})\big), \tag{3.25}$$

with $\lambda^-_{\rho^*_{rs}} := \max_{j\in\mathbb{Z}} \big\{\lambda_j \colon h^{\rho^*_{rs}}(\lambda_j) > 0\big\}$ and $\lambda^+_{\rho^*_{rs}} := \min_{j\in\mathbb{Z}} \big\{\lambda_j \colon h^{\rho^*_{rs}}(\lambda_j) < 0\big\}$. The estimators (3.24) and (3.25) can be shown, under assumptions (C1) and (C3), to satisfy (C2) along the same lines as in Theorem 5.3 of Ilmonen and Paindaveine (2011).

3.3 Data-driven specification of reference density

While the choice of the reference density f has no impact on the consistency properties of the corresponding R-estimator $\underset{\sim}{\mathbf{L}}^{(n)}_f$, it has a direct influence on its performances for both finite n and as $n \to \infty$; the "closer" f is to the actual density g, the better the performance for $\underset{\sim}{\mathbf{L}}^{(n)}_f$. The efficiency loss due to a misspecified reference density f is revealed though an

inspection of the cross-information quantities.

Many mixing matrix estimators of $\mathbf{L}$, including those proposed by Chen and Bickel (2006) and Bach and Jordan (2002), rely on nonparametric estimates of the underlying component densities or scores. However, such nonparametric estimates require large sample sizes to be effective due to their sensitivity to tuning parameters such as bandwidth or choice of basis functions. For instance, Chen and Bickel (2006) propose estimating score functions using a basis of t B-spline functions; the exact choice of t has a significant impact on the resulting estimator. Furthermore, nonparametric methods tend to be sensitive to outliers, especially in the case of small to moderate-sized samples.

The purpose of using the R-estimators based on f-scores is precisely to increase robustness against outliers while avoiding nonparametric density estimation. A distinctive feature of ranks is that they are independent, under the null hypothesis and hence also under contiguous alternatives, of the corresponding order statistics. That property can be exploited, in the spirit of Dodge and Jurečková (2000), to select a reference density f that accounts for features (skewness, kurtosis, etc.) of the actual underlying g: as long as such a selection is based on order statistics, it has no impact on the validity of R-estimation procedures.

We propose selecting $f := (f_1, \ldots, f_k)$ by fitting, componentwise, a parametric density to the (order statistic of the) residuals associated with the preliminary estimator $\tilde{\mathbf{L}}^{(n)}$. If skewness and kurtosis are to be accounted for, a convenient family of densities is the family of skew t-distribution (Azzalini and Capitanio 2003) with densities of the form

$$h_{\boldsymbol{\omega}}(x) = \frac{2}{\sigma} t_\nu(z) T_{\nu+1}\Big(\alpha z \Big(\frac{\nu+1}{\nu+z^2}\Big)^{1/2}\Big) \qquad \text{for } x \in \mathbb{R} \text{ and } z := \sigma^{-1}(x-\mu), \tag{3.26}$$

indexed by $\boldsymbol{\omega} := (\mu, \sigma, \alpha, \nu)$, where $\mu \in \mathbb{R}$ is a location, $\sigma \in \mathbb{R}_0^+$ a scale, $\alpha \in \mathbb{R}$ a skewness parameter, and $\nu > 0$ the number of degrees of freedom governing the tails; $t_\nu(z)$ and $T_\nu(z)$ are the density and cumulative distribution functions, respectively, of Student's t-distribution with ν degrees of freedom. For each $j = 1, \ldots, k$, an estimator $(\hat{\mu}_j, \hat{\sigma}_j, \hat{\alpha}_j, \hat{\nu}_j)$ is obtained

from the residuals $Z_{1,j}^{(n)}(\tilde{\mathbf{L}}^{(n)}), \ldots, Z_{n,j}^{(n)}(\tilde{\mathbf{L}}^{(n)})$ using a method such as maximum likelihood. Then, the f-score functions used in the R-estimation procedure are those associated with the skew t-density $h_{\hat{\boldsymbol{\omega}}_j}$, with $\hat{\boldsymbol{\omega}}_j = (\hat{\mu}_j, \hat{\sigma}_j, \hat{\alpha}_j, \hat{\nu}_j)$, thus taking into account the skewness, kurtosis and tails of the residuals. Data-driven scores, however, clearly need not be restricted to the family of skew t-densities, and can be selected from other univariate parametric families as well; in Section 4, we also consider, for instance, the family of stable distributions, indexed by $\boldsymbol{\omega} := (\mu, \sigma, \beta, \gamma)$, where μ and σ are location and scale, β is a skewness parameter ($\beta = 0$ means symmetry), and $\gamma \in (0, 2]$, the tail index, characterizes the tail behavior ($\gamma = 2$ means Gaussian tails, $\gamma = 1$ Cauchy tails).

4 Simulations

Simulation experiments are conducted to examine finite-sample performances of the proposed R-estimation procedure. In the simulations, we evaluate R-estimators $\underset{\sim}{\mathbf{L}}_f^{(n)}$ based on various preliminary estimators from the literature and a data-driven reference density f, as described in Section 3.3. In this section, we describe the precise construction of the four preliminary estimators to be used, the R-estimator $\underset{\sim}{\mathbf{L}}_f^{(n)}$, and, for the sake of comparison, the R_+-estimator of Ilmonen and Paindaveine (2011). Then we describe the simulation experiment setups and conclude with a discussion of the simulation results.

4.1 Preliminary, R-, and R_+-estimators

4.1.1 The preliminary estimators

Oja et al. (2006) propose estimating a mixing matrix using two distinct *scatter matrices* with the *independent components property*. A *scatter matrix* is a $k \times k$ symmetric positive definite and affine-equivariant function of a sample of n random k-vectors. A scatter matrix is said to possesses the *independent components property* if, when the sample of random k-vectors

at which it is evaluated is i.i.d. with mutually independent components, all of its off-diagonal elements are $o_{\mathrm{P}}(1)$ as the sample size grows to infinity. The sample covariance matrix is a classical example of a scatter matrix exhibiting that property. Generally, however, a scatter matrix possesses the independent components property only if the mutually independent components all possess symmetric distributions, a condition which is not satisfied here.

As a remedy, Nordhausen et al. (2008) propose generalizing Oja et al. (2006) by constructing the estimator using two distinct *symmetrized* scatter matrices. Symmetrizing a scatter matrix $\mathbf{S}(\cdot)$ entails evaluating the same scatter matrix function at the distinct pairwise differences of observations from a given sample. Specifically, if $\mathbf{X}_1^{(n)}, \ldots, \mathbf{X}_n^{(n)}$ is an observed sample of k-vectors, then the symmetrized version of $\mathbf{S}(\cdot)$ is defined to be

$$\mathbf{S}^*(\mathbf{X}_1^{(n)}, \ldots, \mathbf{X}_n^{(n)}) := \mathbf{S}\big(\tilde{\mathbf{X}}_{(1,2)}^{(n)}, \ldots, \tilde{\mathbf{X}}_{(1,n)}^{(n)}, \tilde{\mathbf{X}}_{(2,3)}^{(n)}, \ldots, \tilde{\mathbf{X}}_{(2,n)}^{(n)}, \tilde{\mathbf{X}}_{(3,4)}^{(n)}, \ldots, \tilde{\mathbf{X}}_{(n-1,n)}^{(n)}\big), \quad (4.27)$$

where $\tilde{\mathbf{X}}_{(i,j)}^{(n)} := \mathbf{X}_i^{(n)} - \mathbf{X}_j^{(n)}$ for each $\{(i,j) : \ 1 \le i < j \le n\}$ denotes the $n(n-1)/2$ distinct pairwise differences. If $\mathbf{X}_i^{(n)}$ $(i = 1, \ldots, n)$ is i.i.d. with mutually independent components, then $\tilde{\mathbf{X}}_{(i,j)}^{(n)}$ $(1 \le i < j \le n)$ is also i.i.d. with mutually independent components each having, by construction, symmetric distributions. Consequently, the symmetrized version of any scatter matrix $\mathbf{S}(\cdot)$ has the independent components property.

Letting $\mathbf{S}_{\mathrm{A}}^*$ and $\mathbf{S}_{\mathrm{B}}^*$ denote the symmetrized versions of two distinct scatter matrices $\mathbf{S}_{\mathrm{A}}$ and $\mathbf{S}_{\mathrm{B}}$ as in (4.27) and letting $\mathbf{X}_1^{(n)}, \ldots, \mathbf{X}_n^{(n)}$ denote an observed sample of k-variate mixed data, Nordhausen et al. (2008) propose an estimator $\hat{\mathbf{\Lambda}}\big(\mathbf{S}_{\mathrm{A}}^*, \mathbf{S}_{\mathrm{B}}^*\big)$ that is the $k \times k$ nonsingular matrix $\mathbf{\Lambda}$ simultaneously satisfying

$$\mathbf{S}_{\mathrm{A}}^*\big(\mathbf{\Lambda}^{-1}\mathbf{X}_1^{(n)}, \ldots, \mathbf{\Lambda}^{-1}\mathbf{X}_n^{(n)}\big) = \mathbf{I}_k \quad \text{and} \quad \mathbf{S}_{\mathrm{B}}^*\big(\mathbf{\Lambda}^{-1}\mathbf{X}_1^{(n)}, \ldots, \mathbf{\Lambda}^{-1}\mathbf{X}_n^{(n)}\big) = \mathbf{D}, \quad (4.28)$$

where $\mathbf{D}$ is any full-rank $k \times k$ diagonal matrix. In the simulations below, we construct preliminary estimators based on the following scatter matrices: the sample covariance

$$\mathbf{S}_{\text{COV}} := \frac{1}{n}\sum_{i=1}^{n}\big(\mathbf{X}_i^{(n)} - \bar{\mathbf{X}}^{(n)}\big)\big(\mathbf{X}_i^{(n)} - \bar{\mathbf{X}}^{(n)}\big)' \qquad \text{where} \qquad \bar{\mathbf{X}}^{(n)} := \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i^{(n)};$$

the fourth-order scatter matrix

$$\mathbf{S}_{\text{COV4}} := \frac{1}{n}\sum_{i=1}^{n}\big(\mathbf{X}_i^{(n)} - \bar{\mathbf{X}}^{(n)}\big)'\big(\mathbf{S}_{\text{COV}}\big)^{-1}\big(\mathbf{X}_i^{(n)} - \bar{\mathbf{X}}^{(n)}\big)\big(\mathbf{X}_i^{(n)} - \bar{\mathbf{X}}^{(n)}\big)\big(\mathbf{X}_i^{(n)} - \bar{\mathbf{X}}^{(n)}\big)';$$

and the van der Waerden rank-based estimator $\mathbf{S}_{\text{HOP}}$ (Hallin et al. 2006). Letting $\mathbf{S}^*_{\text{COV}}$, $\mathbf{S}^*_{\text{COV4}}$, and $\mathbf{S}^*_{\text{HOP}}$ denote symmetrized versions of these scatter matrices (4.27), we obtain the preliminary estimators $\tilde{\mathbf{\Lambda}}_{\text{Fobi}} := \tilde{\mathbf{\Lambda}}\big(\mathbf{S}^*_{\text{COV}}, \mathbf{S}^*_{\text{COV4}}\big)$ and $\tilde{\mathbf{\Lambda}}_{\text{HOPCov}} := \tilde{\mathbf{\Lambda}}\big(\mathbf{S}^*_{\text{HOP}}, \mathbf{S}^*_{\text{COV}}\big)$, defined in (4.28), from each sample of k-variate data generated in the simulations. Often referred to as FOBI in the literature, $\tilde{\mathbf{\Lambda}}_{\text{Fobi}}$ was first studied in Cardoso (1989), and is the most usual estimator of that type. As for $\tilde{\mathbf{\Lambda}}_{\text{HOPCov}}$, it can be expected to inherit some of the favorable robustness properties of the rank-based $\mathbf{S}^*_{\text{HOP}}$. When computing $\tilde{\mathbf{\Lambda}}_{\text{Fobi}}$ and $\tilde{\mathbf{\Lambda}}_{\text{HOPCov}}$, obtaining symmetrized scatter matrices introduces a heavy computational burden; for a sample of size n, the symmetrization step requires evaluating a scatter matrix from $n(n-1)/2$ pairwise differences. Because sample sizes in typical ICA applications are large, $\tilde{\mathbf{\Lambda}}_{\text{Fobi}}$ and $\tilde{\mathbf{\Lambda}}_{\text{HOPCov}}$ may be impractical.

The FastICA algorithm (Hyvärinen and Oja (1997); in the simulations, we used the fastICA R package by Marchini et al. (2012), with default settings and the initial demixing matrix set to identity) exploits the canonical assumption in ICA that at most one source component possesses a Gaussian distribution. A mixing matrix is selected by maximizing a contrast function that approximates the sample negentropy sequentially for each component. Because negentropy is a measure of non-Gaussianity, the estimated components obtained via the FastICA estimator $\tilde{\mathbf{\Lambda}}_{\text{FIca}}$ have both low cross-correlation, and all but one possess empirical distributions not well-approximated by a Gaussian. Reyhani et al. (2012) establish sufficient conditions for the root-n consistency of $\tilde{\mathbf{\Lambda}}_{\text{FIca}}$.

Finally, the Kernel-ICA-KGV or Kernel-ICA algorithm (Bach and Jordan 2003) seeks

a demixing matrix that minimizes the mutual information between the implied independent components via *generalized variance*, a construction implicitly measuring non-Gaussianity. Of all preliminary estimators we considered, $\tilde{\mathbf{\Lambda}}_{\text{KIca}}$ (computed from the kernel-ica Matlab package (Bach 2003) with default settings) has the strongest performances in the simulations; its asymptotic properties have not been well studied, though, and conditions for its root-n consistency have not been established.

After evaluating each preliminary estimator ($\tilde{\mathbf{\Lambda}}_{\text{Fobi}}$, $\tilde{\mathbf{\Lambda}}_{\text{HOPCov}}$, $\tilde{\mathbf{\Lambda}}_{\text{FIca}}$, and $\tilde{\mathbf{\Lambda}}_{\text{KIca}}$) from each replication, one-step R-estimators are computed from the observationally equivalent

$$\tilde{\mathbf{L}}_{\text{Fobi}} := \Pi(\tilde{\mathbf{\Lambda}}_{\text{Fobi}}),\ \ \tilde{\mathbf{L}}_{\text{HOPCov}} := \Pi\big(\tilde{\mathbf{\Lambda}}_{\text{HOPCov}}\big),\ \ \tilde{\mathbf{L}}_{\text{FIca}} := \Pi\big(\tilde{\mathbf{\Lambda}}_{\text{FIca}}\big),\ \text{ and }\ \tilde{\mathbf{L}}_{\text{KIca}} := \Pi\big(\tilde{\mathbf{\Lambda}}_{\text{KIca}}\big), \tag{4.29}$$

which belong to $\mathcal{M}_k^1$ (see (1.3) for the definition of the mapping Π).

4.1.2 The R-estimators

As described in Section 3.3, we used data-driven scores from the skew t-family in the construction of our R-estimators. For each replication of $\mathbf{X}_1^{(n)}, \ldots, \mathbf{X}_n^{(n)}$ and preliminary estimator $\tilde{\mathbf{L}} \in \mathcal{M}_k^1$, we compute the residuals $\hat{\mathbf{Z}}_i^{(n)}\big(\tilde{\mathbf{L}}\big) := \tilde{\mathbf{L}}^{-1}\mathbf{X}_i^{(n)}$ for $i = 1, \ldots, n$. For each $j = 1, \ldots, k$, a skew t-density $h_{\hat{\boldsymbol{\omega}}_j}$ (see (3.26)) is fit to the n-tuple $\hat{Z}_{1,j}^{(n)}\big(\tilde{\mathbf{L}}\big), \ldots, \hat{Z}_{n,j}^{(n)}\big(\tilde{\mathbf{L}}\big)$ of jth components via maximum likelihood (MLE). In this implementation, a constrained MLE $\hat{\boldsymbol{\omega}}_j$, with $\hat{\alpha}_j \in [-30, 30]$ and $\hat{\nu}_j \in [3, \infty)$, was adopted for the sake of numerical stability. The resulting one-step R-estimate then is, with $f := (h_{\hat{\boldsymbol{\omega}}_1}, \ldots, h_{\hat{\boldsymbol{\omega}}_k})$,

$$\underset{\sim}{\mathbf{L}}^*\big(\tilde{\mathbf{L}}\big) := \tilde{\mathbf{L}} + n^{-\frac{1}{2}}\ \tilde{\mathbf{L}}\Big[\hat{\mathbf{N}}_{\tilde{\mathbf{L}},f}^{(n)} - \operatorname{diag}\Big(\tilde{\mathbf{L}}\ \hat{\mathbf{N}}_{\tilde{\mathbf{L}},f}^{(n)}\Big)\Big], \tag{4.30}$$

where $\hat{\mathbf{N}}_{\tilde{\mathbf{L}},f}^{(n)}$ is defined in (3.21) (because $\underset{\sim}{\mathbf{L}}^*\big(\tilde{\mathbf{L}}\big)$ is based on data-driven scores, no reference density is used in the notation).

In the simulations, we also explore the performance of a multistep version of the R-estimator just described. Taking $\underset{\sim}{\mathbf{L}}^*\big(\tilde{\mathbf{L}}\big)$ as a preliminary, (4.30) indeed is easily iterated:

letting $\underset{\sim}{\mathbf{L}}^*_{(0)}\big(\tilde{\mathbf{L}}\big) := \tilde{\mathbf{L}}$, define, for $t = 1, \ldots, T$,

$$\underset{\sim}{\mathbf{L}}^*_{(t)}\big(\tilde{\mathbf{L}}\big) := \underset{\sim}{\mathbf{L}}^*\Big(\underset{\sim}{\mathbf{L}}^*_{(t-1)}\big(\tilde{\mathbf{L}}\big)\Big). \tag{4.31}$$

4.1.3 The R_+-estimators

We also computed the signed-rank R_+-estimators proposed by Ilmonen and Paindaveine (2011), the validity of which requires symmetric component densities. The computation of those R_+-estimators not only requires a root-n consistent preliminary estimator $\tilde{\mathbf{L}}^{(n)} \in \mathcal{M}^1_k$, but also an estimate for the location $\boldsymbol{\mu} \in \mathbb{R}^k$. The preliminary estimators we used are those described in Section 4.1; for location, we adopted the same componentwise median estimator as in Ilmonen and Paindaveine (2011). To make the comparison a fair one, however, we also implemented the signed-rank procedure on the basis of data-driven scores, as explained in Section 4.1.2—restricting the fit, of course, to symmetric Student or stable densities. The resulting R_+-estimators are denoted as $\underset{\sim}{\mathbf{L}}^*_+\big(\tilde{\mathbf{L}}\big)$. Finally, parallel to (4.31), multistep versions of $\underset{\sim}{\mathbf{L}}^*_+\big(\tilde{\mathbf{L}}\big)$ are easily constructed; the notation $\underset{\sim}{\mathbf{L}}^*_{+(t)}\big(\tilde{\mathbf{L}}\big)$ is used in an obvious way.

4.2 Simulation experiments

In each simulation experiment, bivariate observations (k=2) were generated from various generating processes. Each generating process is characterized by a sample size n and two component densities, $g_1^{(S)}$ and $g_2^{(S)}$, $S = A, \ldots, L$, the list of which is provided in Table 1, yielding various skewness levels and tail behaviors. We also consider (E) an asymmetric bimodal mixture distribution. Each marginal distribution has median equal to zero (with location parameter set accordingly) and unit scale. The same mixing matrix $\mathbf{L} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \in \mathcal{M}^1_2$ was used throughout; location, which plays no role, was set to $\boldsymbol{\mu} = \mathbf{0}$. Small ($n = 100$) and moderate ($n = 1,000$) sample sizes were considered.

Table 1: Component densities used in the simulation experiment, all with median zero and unit scale: (a) skew $t(\alpha, \nu)$ denotes the skew t-density with shape (asymmetry) parameter α and ν degrees of freedom; (b) stable(β, γ) denotes the stable density with skewness parameter β and tail index γ; (c) asymMix-t_3 is a mixture of two Student t-distributions with 3 degrees of freedom; and (d) t_ν is the classical Student t-distribution with ν degrees of freedom.

	Component densities	
S	$g_1^{(\mathrm{S})}$	$g_2^{(\mathrm{S})}$
(A)	skew $t(\alpha = 5, \nu = 1)$	skew $t(\alpha = 5, \nu = 1)$
(B)	skew $t(\alpha = 5, \nu = 6)$	skew $t(\alpha = 5, \nu = 6)$
(C)	skew $t(\alpha = 5, \nu = 10)$	skew $t(\alpha = 5, \nu = 10)$
(D)	stable$(\beta = 1, \gamma = 1.75)$	skew $t(\alpha = 5, \nu = 10)$
(E)	asymMix-t_3	skew $t(\alpha = 5, \nu = 10)$
(F)	Student's t_{10}	skew $t(\alpha = 5, \nu = 10)$
(G)	stable$(\beta = 1, \gamma = 1.5)$	stable$(\beta = 1, \gamma = 1.5)$
(H)	stable$(\beta = 1, \gamma = 1.75)$	stable$(\beta = 1, \gamma = 1.75)$
(I)	stable$(\beta = 0, \gamma = 1.75)$	stable$(\beta = 0, \gamma = 1.75)$
(J)	Cauchy t_1	Cauchy t_1
(K)	Student's t_6	Student's t_6
(L)	Student's t_{10}	Student's t_{10}

For each generating process (each combination of $n = 100$ or $1,000$ and $S \in \{A, \ldots, L\}$), the number of replications was set to $M = 1,000$, and, for each replication, the following estimators of $\mathbf{L}$ were computed:

(a) the preliminary estimators $\tilde{\mathbf{L}} = \tilde{\mathbf{L}}_{\mathrm{Fobi}}$, $\tilde{\mathbf{L}}_{\mathrm{HOPCov}}$, $\tilde{\mathbf{L}}_{\mathrm{FIca}}$, and $\tilde{\mathbf{L}}_{\mathrm{KIca}}$ given in (4.29);

(b) the one-step R-estimators $\underset{\sim}{\mathbf{L}}^*\big(\tilde{\mathbf{L}}\big)$ based on the preliminary ones as listed under (a) and data-driven skew t-scores;

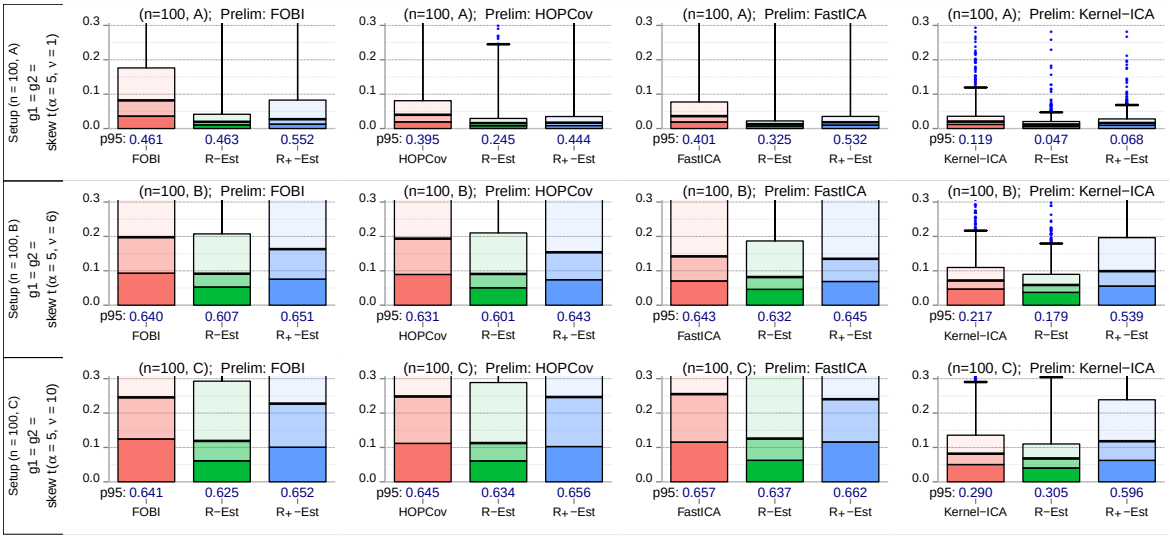
(c) the one-step R_+-estimators $\underset{\sim}{\mathbf{L}}_+^*\big(\tilde{\mathbf{L}}\big)$ based on the preliminary ones as listed under (a) and data-driven Student's t-scores.

For component densities (A), (C) and (H), moreover, we also computed, for $n = 100$ and $n = 1,000$,

(d) the T-multistep versions of the R-estimators based on the preliminary $\tilde{\mathbf{L}}_{\mathrm{HOPCov}}$ and $\tilde{\mathbf{L}}_{\mathrm{KIca}}$, still with data-driven skew t-scores, $T = 1, \ldots, 10$.

Figure 1: Boxplots of Amari errors obtained in $M = 1000$ replications of the setup (n, S), $n = 100,\ 1,000$, $\mathrm{S} = A,\ B,\ C$, for the preliminary $\tilde{\mathbf{L}} = \tilde{\mathbf{L}}_{\mathrm{Fobi}}$, $\tilde{\mathbf{L}}_{\mathrm{HOPCov}}$, $\tilde{\mathbf{L}}_{\mathrm{FIca}}$, $\tilde{\mathbf{L}}_{\mathrm{KIca}}$, the one-step R-estimator $\underset{\sim}{\mathbf{L}}^*(\tilde{\mathbf{L}})$, and the one-step R_+-estimator $\underset{\sim}{\mathbf{L}}^*_+(\tilde{\mathbf{L}})$ based on the same preliminaries with data-driven skew t- and Student's t-scores, respectively.

(a) Sample size $n = 100$



(b) Sample size $n = 1000$

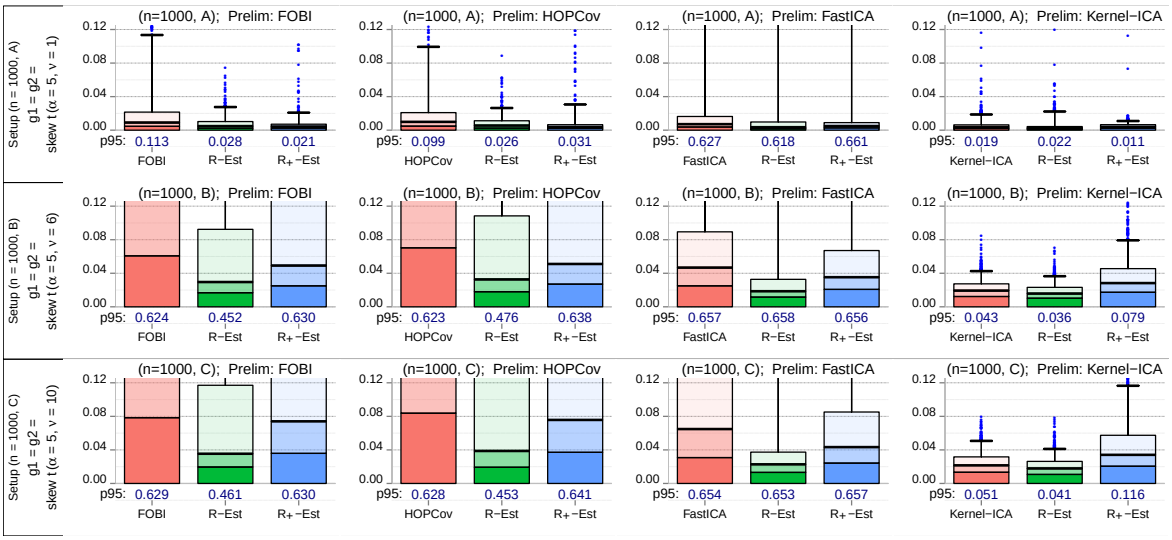
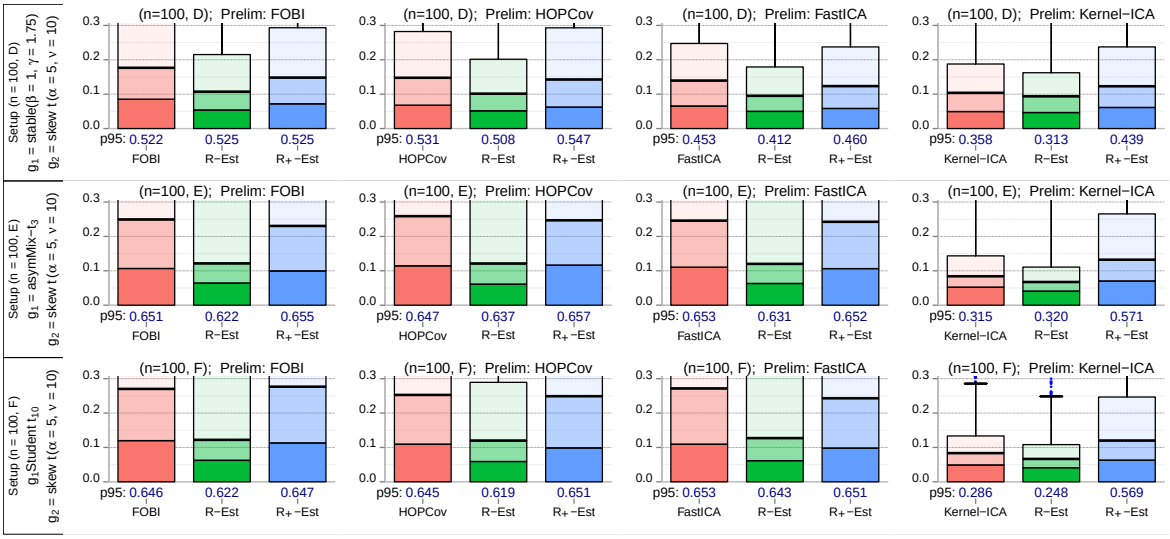
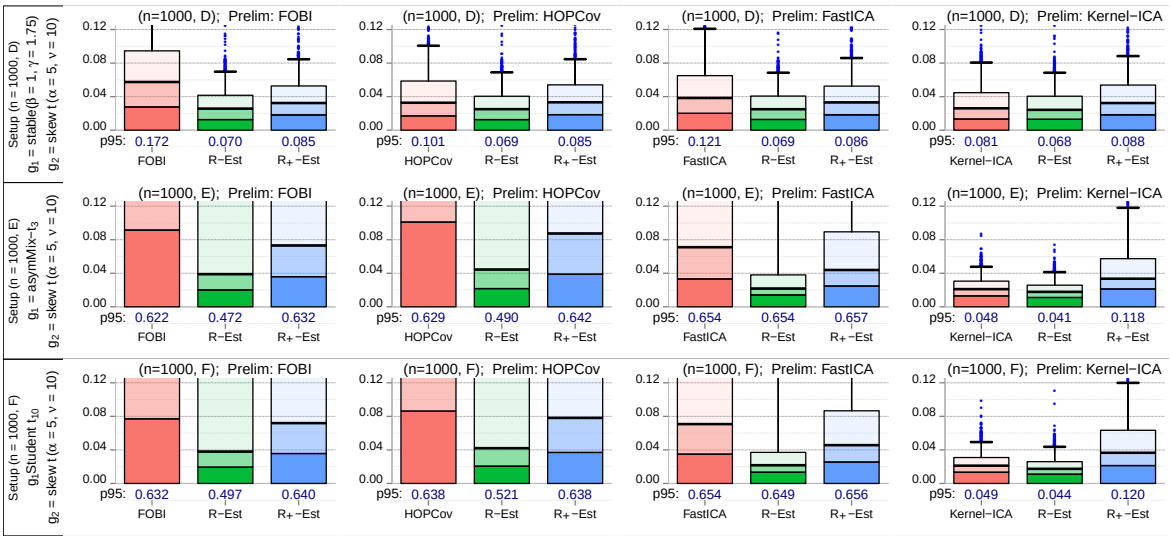


Figure 2: Boxplots of Amari errors obtained in $M = 1000$ replications of the setup (n, S), $n = 100,\ 1,000$, $\mathrm{S} = D,\ E,\ F$, for the preliminary $\tilde{\mathbf{L}} = \tilde{\mathbf{L}}_{\mathrm{Fobi}}$, $\tilde{\mathbf{L}}_{\mathrm{HOPCov}}$, $\tilde{\mathbf{L}}_{\mathrm{FIca}}$, $\tilde{\mathbf{L}}_{\mathrm{KIca}}$, the one-step R-estimator $\underset{\sim}{\mathbf{L}}^*(\tilde{\mathbf{L}})$, and the one-step R_+-estimator $\underset{\sim}{\mathbf{L}}^*_+(\tilde{\mathbf{L}})$ based on the same preliminaries with data-driven skew t- and Student's t-scores, respectively.

(a) Sample size $n = 100$



(b) Sample size $n = 1000$

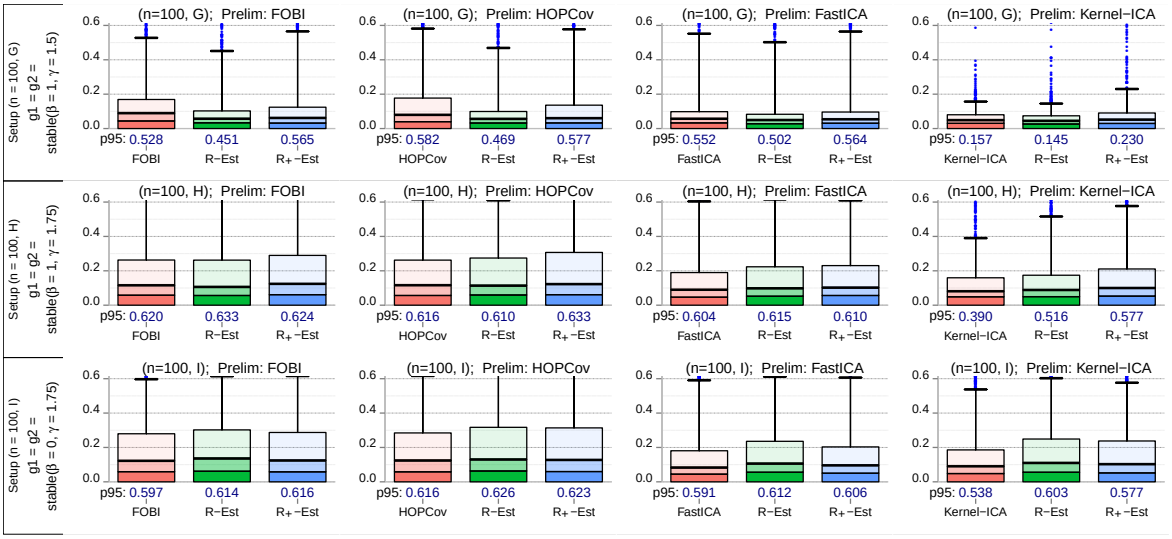


The performance of each estimator ($\tilde{\mathbf{L}}$, $\underset{\sim}{\mathbf{L}}^*$ or $\underset{\sim}{\mathbf{L}}^*_+$) is measured by its *Amari error* with respect to $\mathbf{L}$. The Amari error (Amari et al. 1996) $\mathrm{AE}(\mathbf{A}, \mathbf{B})$ of a $k \times k$ matrix $\mathbf{A}$ with respect to a nonsingular $k \times k$ matrix $\mathbf{B}$ is defined as

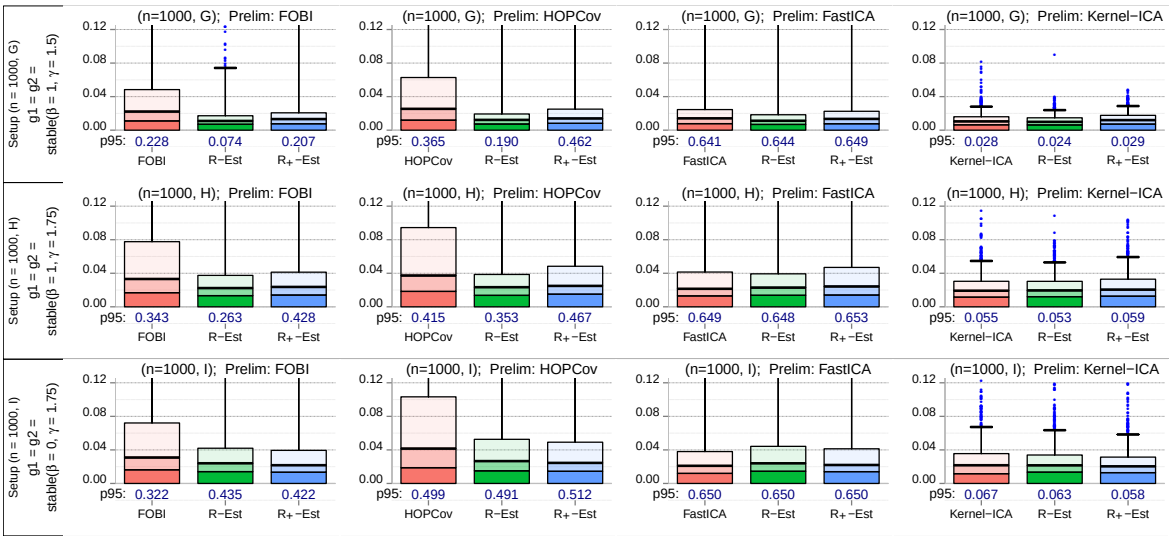
$$\mathrm{AE}(\mathbf{A}, \mathbf{B}) = \frac{1}{2k(k-1)} \left(\sum_{i=1}^{k} \left(\frac{\sum_{j=1}^{k} |w_{ij}|}{\max_j |w_{ij}|} - 1 \right) + \sum_{j=1}^{k} \left(\frac{\sum_{i=1}^{k} |w_{ij}|}{\max_i |w_{ij}|} - 1 \right) \right), \tag{4.32}$$

Figure 3: Boxplots of Amari errors obtained in $M = 1000$ replications of the setup (n, S), $n = 100,\ 1,000$, $\mathrm{S} = G,\ H,\ I$, for the preliminary $\tilde{\mathbf{L}} = \tilde{\mathbf{L}}_{\mathrm{Fobi}}$, $\tilde{\mathbf{L}}_{\mathrm{HOPCov}}$, $\tilde{\mathbf{L}}_{\mathrm{FIca}}$, $\tilde{\mathbf{L}}_{\mathrm{KIca}}$, the one-step R-estimator $\underset{\sim}{\mathbf{L}}^*(\tilde{\mathbf{L}})$, and the one-step R_+-estimator $\underset{\sim}{\mathbf{L}}^*_+(\tilde{\mathbf{L}})$ based on the same preliminaries with data-driven skew t- and Student's t-scores, respectively.

(a) Sample size $n = 100$



(b) Sample size $n = 1000$

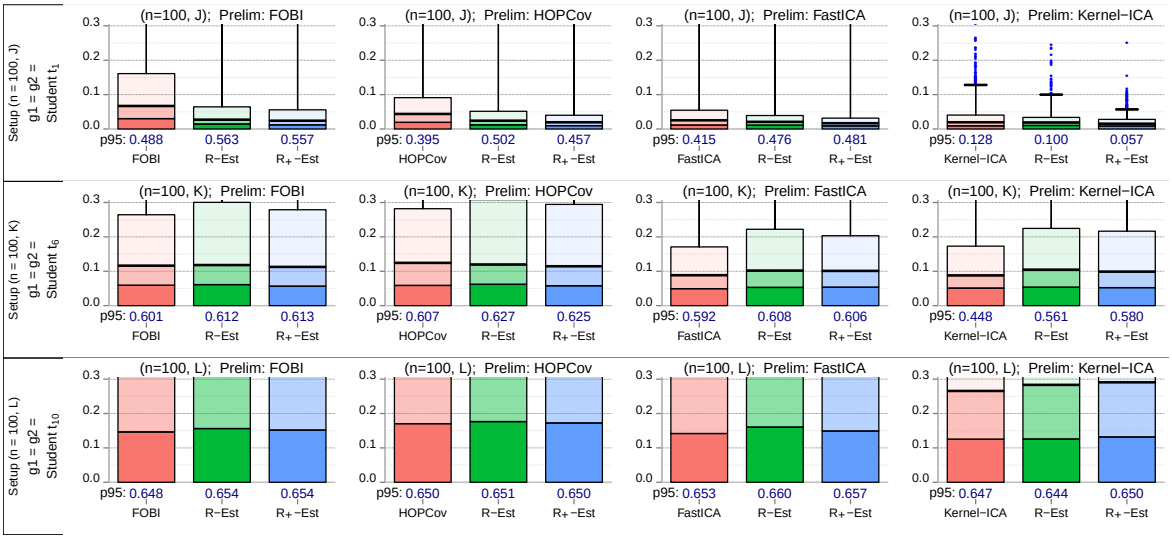


with $\mathbf{W} := \mathbf{B}^{-1}\mathbf{A} = [w_{ij}]$. The Amari error (which is not a matrix norm) takes values between 0 and 1; $\mathrm{AE}(\mathbf{A}, \mathbf{B})$ close to 0 indicates higher similarity between $\mathbf{A}$ and $\mathbf{B}$. The value of $\mathrm{AE}(\mathbf{A}, \mathbf{B})$ is invariant under permutations and (positive) rescaling of rows or columns of $\mathbf{A}$ and $\mathbf{B}$. More precisely, $\mathrm{AE}(\mathbf{A}, \mathbf{B}) = \mathrm{AE}(\mathbf{A}^*, \mathbf{B}^*)$ for any $k \times k$ matrices $\mathbf{A}^* := \mathbf{C}_1\mathbf{A}\mathbf{C}_2$ and $\mathbf{B}^* := \mathbf{C}_3\mathbf{B}\mathbf{C}_4$ so long as each $\mathbf{C}_j$ is an arbitrary product of permutation matrices

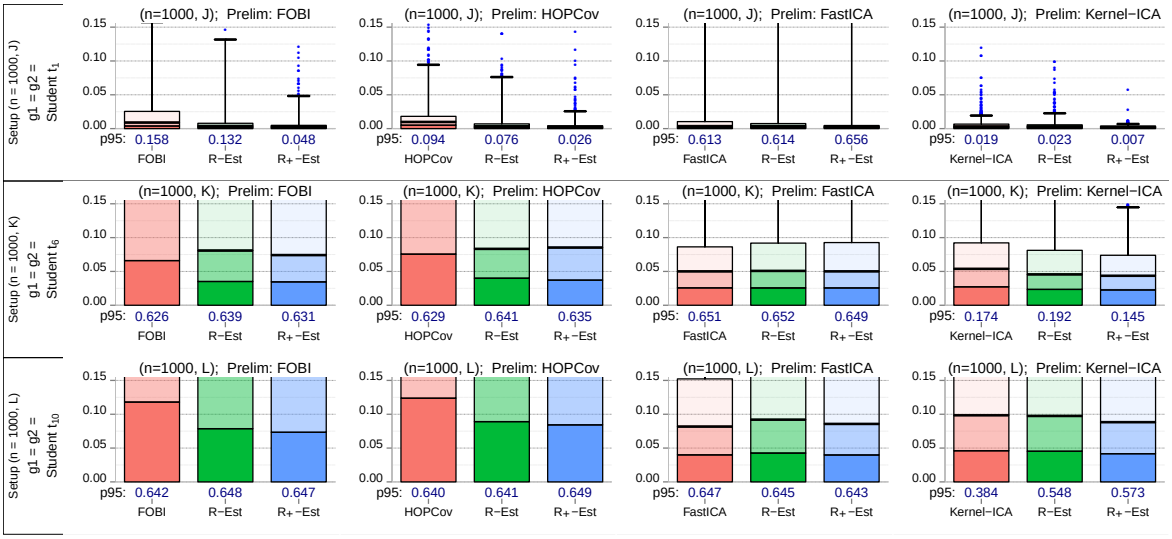
and diagonal matrices with all diagonal entries being positive. Hence, $\mathrm{AE}(\mathbf{A}, \mathbf{B}) = 0$ implies $\Pi(\mathbf{A}) = \Pi(\mathbf{B})$, i.e. $\mathbf{A}$ and $\mathbf{B}$ are observationally equivalent, see (1.3). Therefore, the Amari error is a natural measure of performance for estimators of mixing matrices in ICA.

Figure 4: Boxplots of Amari errors obtained in $M = 1000$ replications of the setup (n, S), $n = 100,\ 1,000$, $\mathrm{S} = J,\ K,\ L$, for the preliminary $\tilde{\mathbf{L}} = \tilde{\mathbf{L}}_{\mathrm{Fobi}}$, $\tilde{\mathbf{L}}_{\mathrm{HOPCov}}$, $\tilde{\mathbf{L}}_{\mathrm{FIca}}$, $\tilde{\mathbf{L}}_{\mathrm{KIca}}$, the one-step R-estimator $\underset{\sim}{\mathbf{L}}^*(\tilde{\mathbf{L}})$, and the one-step R_+-estimator $\underset{\sim}{\mathbf{L}}^*_+(\tilde{\mathbf{L}})$ based on the same preliminaries with data-driven skew t- and Student's t-scores, respectively.

(a) Sample size $n = 100$



(b) Sample size $n = 1000$



Figures 1-6 below are providing boxplots for the $M = 1,000$ Amari distances associated with the various simulation setups. Since Amari distances are intrinsically nonnegative, these

are "one-sided boxplots", showing the first quartile, the median, the third quartile, and a 0.95 quantile whisker. Figures 1-4 are dealing with components densities (A)-(B)-(C), (D)-(E)-(F), (G)-(H)-(I), and (J)-(K)-(L), respectively. Figures 5-6 show the results for the T-step versions of the R-estimators based on $\tilde{\mathbf{L}}_{\mathrm{HOPCov}}$ and $\tilde{\mathbf{L}}_{\mathrm{KIca}}$, under components densities (A)-(C)-(H), as described in (d) above.

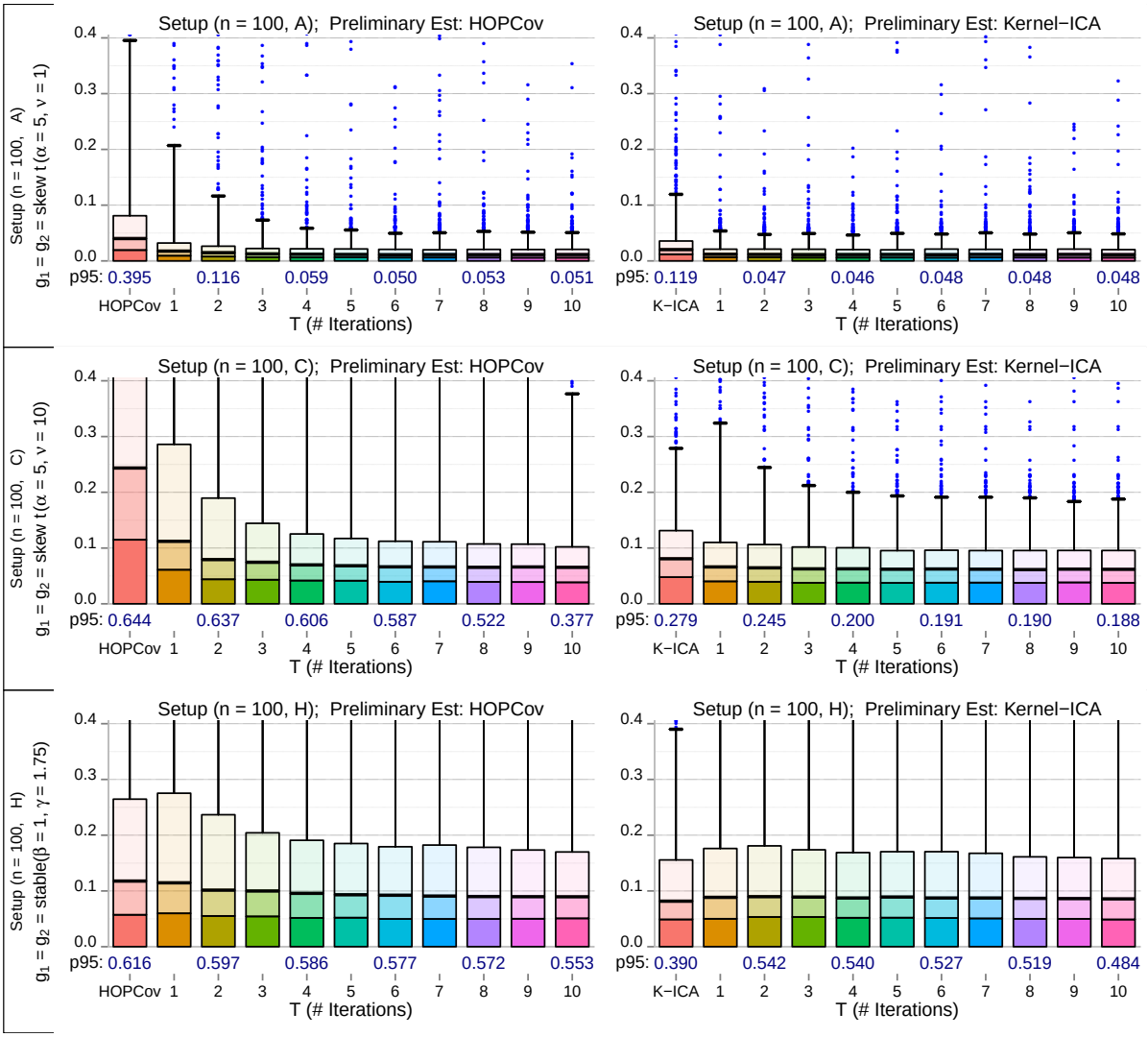


Figure 5: Boxplots of Amari errors obtained in $M = 1000$ replications of the setup (n, S), $n = 100$, $\mathrm{S} = A$, C, H, for the T-step R-estimator $\underset{\sim}{\mathbf{L}}^*(\tilde{\mathbf{L}})$ based on preliminary $\tilde{\mathbf{L}} = \tilde{\mathbf{L}}_{\mathrm{HOPCov}}$ and $\tilde{\mathbf{L}}_{\mathrm{KIca}}$, respectively, and data-driven skew t-scores, $T = 1, \ldots, 10$.

Inspection of Figures 1-4 reveals that Kernel-ICA is, almost uniformly, and sometimes quite substantially (see Figure 1, $n = 1,000$ with 6- and 10-degress of freedom skew t-component densities, or Figure 3, $n = 1,000$ under settings (H) and (I)), the best preliminary.

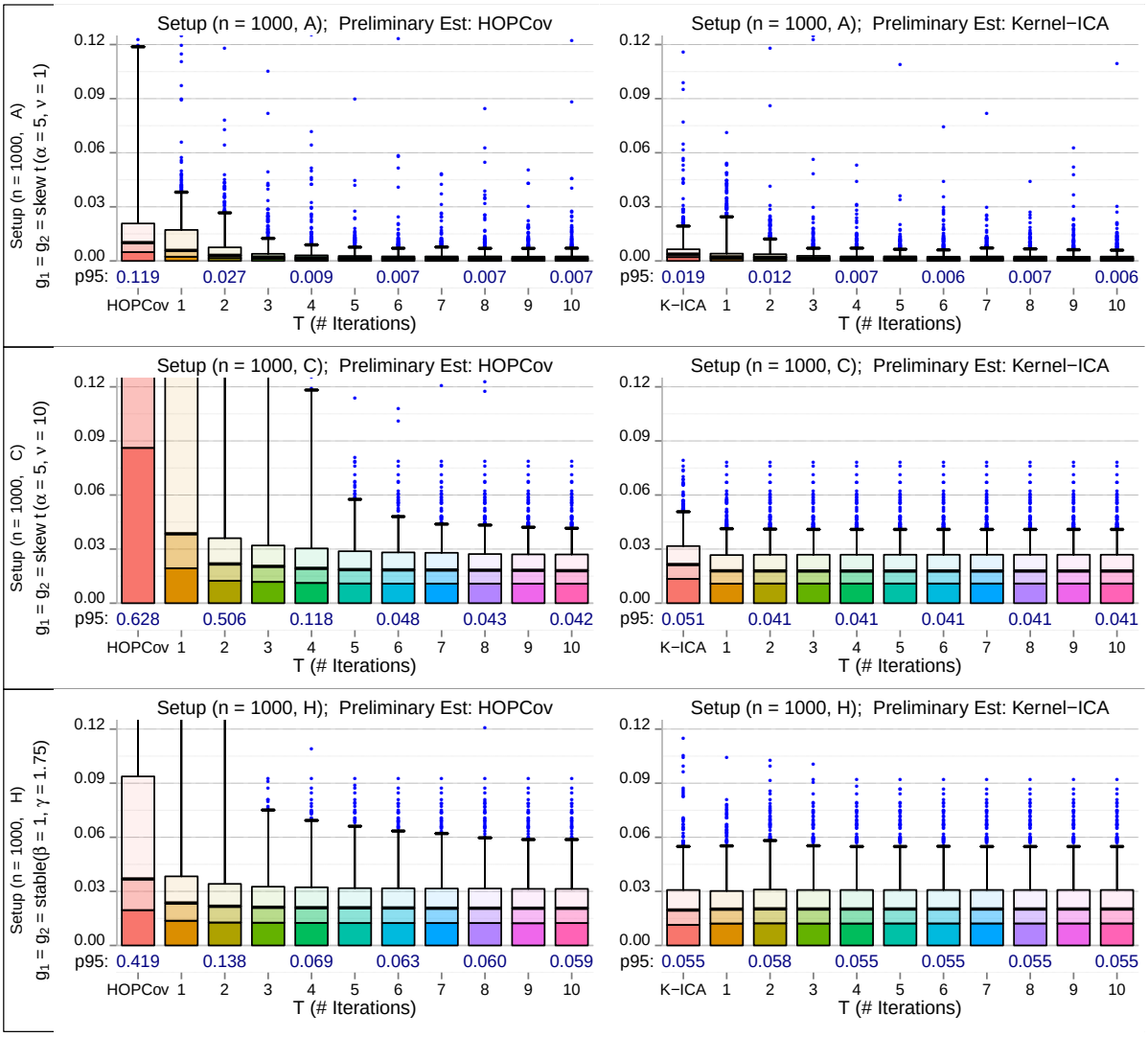


Figure 6: Boxplots of Amari errors obtained in $M = 1000$ replications of the setup (n, S), $n = 1,000$, $\mathrm{S} = A$, C, H, for the T-step R-estimator $\underset{\sim}{\mathbf{L}}^*(\tilde{\mathbf{L}})$ based on preliminary $\tilde{\mathbf{L}} = \tilde{\mathbf{L}}_{\mathrm{HOPCov}}$ and $\tilde{\mathbf{L}}_{\mathrm{KIca}}$, respectively, and data-driven skew t-scores, $T = 1, \ldots, 10$.

Combined with R-estimation (data driven skew t-scores), they are the typical winners, even under symmetric component densities where, in principle, R_+-estimators should do better. The best performances of R-estimators seem to take place under heavy tails (Cauchy and stable component densities)—thanks, probably, to the data-driven selection of scores. Based on FastICA or Kernel-ICA preliminaries, R-estimators moreover are the only ones providing reasonably good results under the "mixed cases" of Figure 2 (bimodal/unimodal, stable/skew t_{10}-, symmetric/skew t- component densities); note that partial symmetry (setup (F)) does not really help R_+-estimation much.

Figures 5 and 6 shows how iterating the rank-based correction can improve a poor preliminary. The HOPCov estimator is typically outperformed by the Kernel-ICA one; however, after a few iteration, both the HOPCov- and Kernel-ICA-based R-estimator are performing quite similarly; the latter, however, needs less iterations than the former to reach its best performance. For $n = 1,000$, starting from Kernel-ICA, one step is essentially sufficient.

5 An application in image analysis

The objective of ICA in applications is typically to recover source signals from a sequence of observed mixed signals. As such, they are widely used in a variety of contexts where the fundamental assumptions (1.1)-(1.2) of ICA are unlikely to hold. One of the merits of existing ICA such as FastICA and Kernel-ICA is that they resist reasonably well to such theoretically unwarranted applications. Such statements, of course, remain unavoidably vague: in the absence of a formal model, indeed, pertinent benchmarks for performance evaluation are hard to define. Demixing acoustic signals or images, where "readability" of the final result appears as an obvious criterion, are an exception. Therefore, in this section, we apply various ICA estimation methods, including the rank-based ones, to the demixing of images that clearly do not satisfy the assumptions we have been making throughout this paper. The results are shown in Figure 7. Their quality is best evaluated by eye-inspection, but a quantitative assessment can be made via the Amari distances provided in Table 8a and b. Although traditional ICA techniques provide reasonable results, our rank-based techniques appear to bring quite significant improvements.

A black-and-white digital image with resolution $h \times w$ ($h, w \in \mathbb{N}$) can be represented by a *pixel matrix* $\mathbf{Z} = (Z_{rs}) \in [0,1]^{h \times w}$, where Z_{rs} represents the "greyness" of the pixel located in the rth row and sth column; if $Z_{rs} = 0$, the pixel is pure black, and if $Z_{rs} = 1$, the pixel is pure white. In this example, we mix three source images of US currency notes, represented by the pixel matrices $\mathbf{Z}_j = (Z_{j;rs})$, $j = 1, 2, 3$ ($h := 65$ and $w := 150$). These three

source images are turned into three mixed ones, with pixel matrices $\mathbf{X}_j = (X_{j;rs})$, $j = 1, 2, 3$, where $(X_{1;rs}, X_{2;s}, X_{3;rs})' = \mathbf{L}^\star(Z_{1;rs}, Z_{2;s}, Z_{3;rs})'$, with $\mathbf{L}^\star = \mathbf{I}_3 + 0.95(\mathbf{1}_3 - \mathbf{I}_3) \in \mathcal{M}_3^1$ (denoting by $\mathbf{1}_3$ a 3×3 matrix of ones); $\mathbf{L}^\star$ thus has a diagonal of ones, all off-diagonal enties being 0.95. The source and mixed images are displayed in Figure 7a.

We then performed ICA estimation on the $n = 65 \times 150 = 9,750$ three-dimensional observations $(X_{1;rs}, X_{2;s}, X_{3;rs})$ by computing the multistep R-estimators $\underset{\sim}{\mathbf{L}}^*_{(T)}(\tilde{\mathbf{L}})$ with data-driven skew t-scores (4.31) and preliminary estimators $\tilde{\mathbf{L}} = \tilde{\mathbf{L}}_{\text{Fobi}}$, $\tilde{\mathbf{L}}_{\text{FIca}}$, and $\tilde{\mathbf{L}}_{\text{KIca}}$ as described in (4.29), and $T = 1, \ldots, 20$; the $\tilde{\mathbf{L}}_{\text{HOPCov}}$ preliminary was omitted because symmetrizing the *HOP* scatter matrix (about 10^8 pairwise differences) was computationally too heavy. Figures 7b, 7c, and 7d contain the resulting $\tilde{\mathbf{L}}$- and $\underset{\sim}{\mathbf{L}}^*_{(20)}(\tilde{\mathbf{L}})$-demixed images. Of all preliminary estimators considered, $\tilde{\mathbf{L}}_{\text{KIca}}$ seems to provide the best results. In Figure 7d, we therefore also provide the demixed images resulting from the Ilmonen and Paindaveine estimator $\underset{\sim}{\mathbf{L}}^*_{+(T)}\big(\tilde{\mathbf{L}}_{\text{KIca}}\big)$ with kernel-ICA preliminary. Irrespective of the preliminary, there is a clear and quite significant visual enhancement, attributable to the use of ranks, in the R-estimation method. Our R-estimators, moreover, substantially outperform the signed-rank ones.

Those eye-inspection conclusions are confirmed and reinforced by the graphs in Figure 8, which reports the Amari errors $\text{AE}\big(\underset{\sim}{\mathbf{L}}^*_{(T)}(\tilde{\mathbf{L}}), \mathbf{L}^\star\big)$ (4.32) for the R- and R_+-estimators of $\mathbf{L}^\star$ and $T = 0, \ldots, 20$. As T increases, for all multistep R-estimators those errors appear to converge to some common limit independent of the preliminary $\tilde{\mathbf{L}}$. For $\tilde{\mathbf{L}} = \tilde{\mathbf{L}}_{\text{FIca}}$ or $\tilde{\mathbf{L}}_{\text{KIca}}$, the decrease is quite significant over $T = 1, \ldots, 5$. The same decrease is much slower for $\tilde{\mathbf{L}} = \tilde{\mathbf{L}}_{\text{Fobi}}$, but the final result, as T gets close to 20, is the same, suggesting that rank-based corrections eventually compensate for a poorer performance of the preliminary estimator.

The same Amari errors $\text{AE}\big(\underset{\sim}{\mathbf{L}}^*_{+(T)}\big(\tilde{\mathbf{L}}\big)), \mathbf{L}^\star\big)$ were evaluated for the multistep (and data-driven-score) versions $\underset{\sim}{\mathbf{L}}^*_{+(T)}\big(\tilde{\mathbf{L}}\big)$ of the Ilmonen and Paindaveine R_+-estimators. The results,

Figure 7: Figure 7a contains the three source images and the three mixed ones. Figures 7b, 7c, and 7d show the demixed images obtained from multistep data-driven skew t-score R-estimators, based on FOBI, FastICA, and Kernel-ICA preliminaries, respectively. In Figure 7d, the result of a Kernel-ICA-based, data-driven Student's t-score multistep R_+-estimator method are also provided.

(a) Top row: the three source images. Bottom row: the three mixed images.

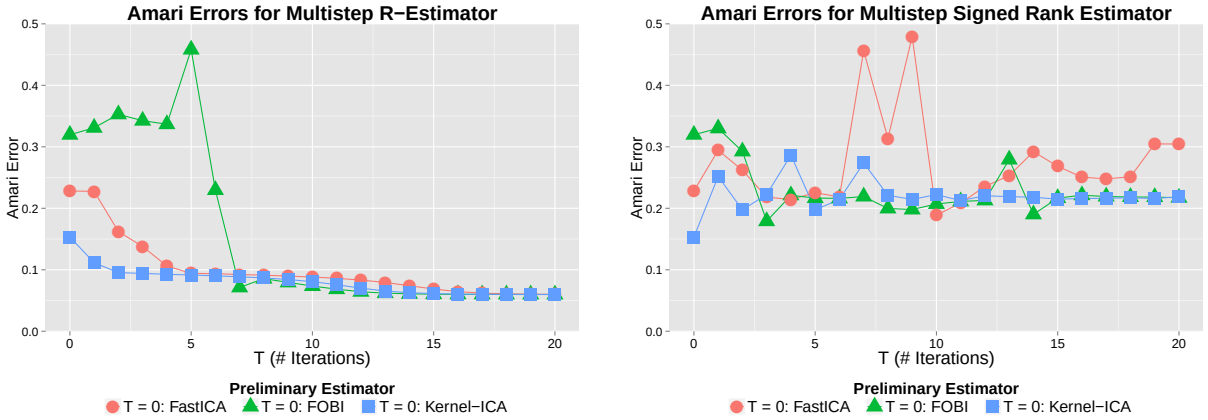
(b) FOBI preliminary. Top row: the $\tilde{\mathbf{L}}_{\text{Fobi}}$-demixed images. Bottom row: the $\underset{\sim}{\mathbf{L}}^*_{(20)}(\tilde{\mathbf{L}}_{\text{Fobi}})$-demixed images.

(c) FastICA preliminary. Top row: the $\tilde{\mathbf{L}}_{\text{FIca}}$-demixed images. Bottom row: the $\underset{\sim}{\mathbf{L}}^*_{(20)}(\tilde{\mathbf{L}}_{\text{FIca}})$-demixed images.

(d) Kernel-ICA preliminary. Top row: the $\tilde{\mathbf{L}}_{\text{KIca}}$-demixed images. Middle row: the $\underset{\sim}{\mathbf{L}}^*_{(20)}(\tilde{\mathbf{L}}_{\text{KIca}})$-demixed images. Bottom row: the $\underset{\sim}{\mathbf{L}}^*_{+(20)}(\tilde{\mathbf{L}}_{\text{KIca}})$-demixed images.

in Figure 8b, clearly show that the signed-rank method fails, which is hardly surprising, since there is little reason for "greyness" in the source images considered here to exhibit any symmetric behavior.

Figure 8: The Amari errors $\mathrm{AE}\big(\hat{\mathbf{L}}, \mathbf{L}^\star\big)$ for the multistep R-estimators $\underset{\sim}{\mathbf{L}}^*_{(T)}(\tilde{\mathbf{L}})$ and the multistep R_+-estimators $\underset{\sim}{\mathbf{L}}^*_{+(T)}(\tilde{\mathbf{L}})$ shown in Figure 7 and based on the preliminary estimators $\tilde{\mathbf{L}} = \tilde{\mathbf{L}}_{\mathrm{Fobi}}$, $\tilde{\mathbf{L}}_{\mathrm{FIca}}$, and $\tilde{\mathbf{L}}_{\mathrm{KIca}}$, for $T = 1, \ldots, 20$.



(a) Multistep R-estimators.

(b) Multistep R_+-estimators.

References

Amari, S., Cichocki, A., Yang, H., et al. (1996). A new learning algorithm for blind signal separation. *Advances in neural information processing systems*, pages 757–763.

Bach, F. (2003). Kernel ica matlab code. `http://www.di.ens.fr/~fbach/kernel-ica/index.htm`. kernel-ica version 1.2, Accessed: Feb. 2, 2012.

Bach, F. and Jordan, M. (2003). Kernel independent component analysis. *The Journal of Machine Learning Research*, 3:1–48.

Bickel, P. J., Klaassen, C., Ritov, Y., and Wellner, J. (1993). *Efficient and adaptive inference in semiparametric models*. Johns Hopkins University Press, Baltimore.

Breiman, L. (1992). Probability, classics in applied mathematics, vol. 7. *Society for Industrial and Applied Mathematics (SIAM), Pennsylvania*.

Cardoso, J. (1989). Source separation using higher order moments. In *Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on*, pages 2109–2112. IEEE.

Cassart, D., Hallin, M., and Paindaveine, D. (2010). On the estimation of cross-information quantities in rank-based inference. *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková*, pages 35–45.

Chen, A. and Bickel, P. (2005). Consistent independent component analysis and prewhitening. *Signal Processing, IEEE Transactions on*, 53(10):3625–3632.

Chen, A. and Bickel, P. (2006). Efficient independent component analysis. *The Annals of Statistics*, 34(6):2825–2855.

Dodge, Y. and Jurečková, J. (2000). *Adaptive regression*. Springer Verlag.

Hájek, J., Šidák, Z., and Sen, P. (1967). *Theory of rank tests*. Academic press New York.

Hallin, M. and Mehta, C. (2013). Sign-and-rank tests in multiple output regression with ica errors. Unpublished manuscript.

Hallin, M., Oja, H., and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. ii. optimal r-estimation of shape. *The Annals of Statistics*, 34(6):2757–2789.

Hallin, M. and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. i. optimal rank-based tests for sphericity. *The Annals of Statistics*, 34(6):2707–2756.

Hallin, M. and Paindaveine, D. (2013). Estimation of cross-information quantities. Unpublished manuscript.

Hallin, M. and Werker, B. (2003). Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli*, 9(1):137–165.

Hyvärinen, A. and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural computation*, 9(7):1483–1492.

Ilmonen, P. and Paindaveine, D. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *The Annals of Statistics*, 39(5):2448–2476.

Kreiss, J. (1987). On adaptive estimation in stationary arma processes. *The Annals of Statistics*, pages 112–133.

Marchini, J. L., Heaton, C., and Ripley, B. D. (2012). *fastICA: FastICA Algorithms to perform ICA and Projection Pursuit*. R package version 1.1-16.

Nordhausen, K., Oja, H., and Ollila, E. (2008). Robust independent component analysis based on two scatter matrices. *Austrian Journal of Statistics*, 37(1):91–100.

Oja, H., Sirkiä, S., and Eriksson, J. (2006). Scatter matrices and independent component analysis. *Austrian Journal of Statistics*, 35(2):175–189.

Reyhani, N., Ylipaavalniemi, J., Vigário, R., and Oja, E. (2012). Consistency and asymptotic normality of fastica and bootstrap fastica. *Signal Processing*, 92(8):1767–1778.

A Appendix: Proofs

A.1 Proof of Proposition 2.1

Oja et al. (2010) establish ULAN for ICA models under the assumption that each f_j is symmetric. Their proof consists in showing that the sufficient conditions of Lemma 1 in Swensen (1985) are satisfied. Mutatis mutandis, that proof still goes through in the present case, with the same central sequence; only the information matrix is affected. That matrix depends on the covariance of $\mathrm{vec}\big(\mathbf{T}^{(n)}_{\mathbf{L};\boldsymbol{\mu},f}\big)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L};f}$, which takes the form

$$\mathrm{E}\big[\mathrm{vec}\big(\mathbf{T}^{(n)}_{\mathbf{L};\boldsymbol{\mu},f}\big)\mathrm{vec}\big(\mathbf{T}^{(n)}_{\mathbf{L};\boldsymbol{\mu},f}\big)'\big] = \sum_{r,s,p,q=1}^{k} \mathrm{E}\big[\big(\mathbf{T}^{(n)}_{\mathbf{L};\boldsymbol{\mu},f}\big)_{r,p}\big(\mathbf{T}^{(n)}_{\mathbf{L};\boldsymbol{\mu},f}\big)_{s,q}\big]\mathbf{e}_p\mathbf{e}_q'\otimes\mathbf{e}_r\mathbf{e}_s'.$$

Because $\big(\mathbf{T}^{(n)}_{\boldsymbol{\mu},\mathbf{L},f}\big)_{r,p}$ is a sum of i.i.d. random variables with expectation zero,

$$\mathrm{E}\big[\big(\mathbf{T}^{(n)}_{\mathbf{L};\boldsymbol{\mu},f}\big)_{r,p}\big(\mathbf{T}^{(n)}_{\mathbf{L};\boldsymbol{\mu},f}\big)_{s,q}\big] = \mathrm{E}\big[(\varphi_{f_r}(Z_{1,r})Z_{1,p}-\delta_{rp})(\varphi_{f_s}(Z_{1,s})Z_{1,q}-\delta_{sq})\big] \quad r,s,p,q\in\{1,\ldots,k\}$$

where the $Z_{1,j}$'s are i.i.d. with density f_j under $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L};f}$ and δ_{rp} is the classical Kronecker index. Evaluating those expectations yields $\mathbf{G}_f$ defined in (2.5). □

A.2 Proofs for Propositions 2.2 and 2.3

Propositions 2.2(i) and 2.3(i) follow from Lemma A.1 below, itself adapted from Theorem V.1.8 in Hájek and Šidák (1967). Consider a triangular array $\big(U^{(n)}_1,V^{(n)}_1\big),\ldots,\big(U^{(n)}_n,V^{(n)}_n\big)$, $n\in\mathbb{N}$ and two scores φ_U, φ_V such that

(D1) $U^{(n)}_i$ and $V^{(n)}_i$, $i=1,\ldots,n$, are uniform over $[0,1]$ and mutually independent, and

(D2) φ_U, $\varphi_V:(0,1)\to\mathbb{R}$ are square-integrable and satisfy (A5).

Denote by $R_i^{(n)}$ the rank of $U_i^{(n)}$ amongst $U_i^{(n)},\dots,U_i^{(n)}$, by $Q_i^{(n)}$ the rank of $V_i^{(n)}$ amongst $V_1^{(n)},\dots,V_n^{(n)}$, and define

$$a_{\mathrm{ex}}^{(n)}(i) := \mathrm{E}\big[\varphi_U(U_1^{(n)})|R_1^{(n)}=i\big], \qquad a_{\mathrm{appr}}^{(n)}(i) := \varphi_U\big(\frac{i}{n+1}\big),$$
$$b_{\mathrm{ex}}^{(n)}(i) := \mathrm{E}\big[\varphi_V(V_1^{(n)})|Q_1^{(n)}=i\big], \qquad \text{and} \qquad b_{\mathrm{appr}}^{(n)}(i) := \varphi_V\big(\frac{i}{n+1}\big).$$

Assumption (D2) implies

$$\lim_{n\to\infty}\frac{\sum_{i=1}^n\big(a_{\mathrm{appr}}^{(n)}(i)-\overline{a}^{(n)}\big)^2}{\max_{1\le i\le n}\big(a_{\mathrm{appr}}^{(n)}(i)-\overline{a}^{(n)}\big)^2}=\infty \quad\text{and}\quad \lim_{n\to\infty}\frac{\sum_{i=1}^n\big(b_{\mathrm{ex}}^{(n)}(i)-\overline{\varphi}_V\big)^2}{\max_{1\le i\le n}\big(b_{\mathrm{ex}}^{(n)}(i)-\overline{\varphi}_V\big)^2}=\infty. \quad (A.33)$$

Let

$$S_{\mathrm{ex}}^{(n)} := \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(a_{\mathrm{ex}}^{(n)}\big(R_i^{(n)}\big)b_{\mathrm{ex}}^{(n)}\big(Q_i^{(n)}\big)-\bar{\varphi_U}\bar{\varphi_V}\Big), \quad (A.34)$$

where $\bar{\varphi_U} := \int_0^1\varphi_U(u)\mathrm{d}u$ and $\bar{\varphi_V} := \int_0^1\varphi_V(v)\mathrm{d}v$; note that

$$\overline{\varphi}_U = \mathrm{E}\big[\varphi_U(U_1^{(n)})\big] = \frac{1}{n}\sum_{i=1}^n\mathrm{E}\big[\varphi_U(U_1^{(n)})|R_1^{(n)}=i\big] = \frac{1}{n}\sum_{i=1}^n a_{\mathrm{ex}}^{(n)}\big(i\big)$$

and, similarly, $\overline{\varphi}_V = \frac{1}{n}\sum_{i=1}^n b_{\mathrm{ex}}^{(n)}(i)$. Also define

$$S_{\mathrm{appr}}^{(n)} := \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(a_{\mathrm{appr}}^{(n)}\big(R_i^{(n)}\big)b_{\mathrm{appr}}^{(n)}\big(Q_i^{(n)}\big)-\overline{a}_{\mathrm{appr}}^{(n)}\overline{b}_{\mathrm{appr}}^{(n)}\Big), \quad (A.35)$$

where $\overline{a}_{\mathrm{appr}}^{(n)} := \frac{1}{n}\sum_{i=1}^n a_{\mathrm{appr}}^{(n)}\big(i\big)$ and $\overline{b}_{\mathrm{appr}}^{(n)} := \frac{1}{n}\sum_{i=1}^n b_{\mathrm{appr}}^{(n)}\big(i\big)$. The following Lemma shows that both $S_{\mathrm{ex}}^{(n)}$ and $S_{\mathrm{appr}}^{(n)}$ admit the asymptotic representation

$$T^{(n)} := \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(\varphi_U\big(U_i^{(n)}\big)\varphi_V\big(V_i^{(n)}\big)-\overline{\varphi}_U^{(n)}\overline{\varphi}_V^{(n)}\Big), \quad (A.36)$$

where $\overline{\varphi}_U^{(n)} = \frac{1}{n}\sum_{i=1}^n\varphi_U\big(U_i^{(n)}\big)$ and $\overline{\varphi}_V^{(n)} = \frac{1}{n}\sum_{i=1}^n\varphi_V\big(V_i^{(n)}\big)$.

Lemma A.1. *Let $\big(U_1^{(n)},V_1^{(n)}\big),\dots,\big(U_n^{(n)},V_n^{(n)}\big)$ and the scores φ_U, φ_V satisfy (D1)-(D2).*

Then, as $n \to \infty$*,*

$$(i)\quad S_{\text{appr}}^{(n)} = S_{\text{ex}}^{(n)} + o_{L^2}(1) \qquad \textit{and} \qquad (ii)\quad S_{\text{appr}}^{(n)} = T^{(n)} + o_{L^2}(1), \tag{A.37}$$

with $S_{\text{ex}}^{(n)}$*,* $S_{\text{appr}}^{(n)}$*, and* $T^{(n)}$ *defined in (A.34), (A.35), and (A.36), respectively.*

Proof. Let us show that

$$(i')\quad \lim_{n\to\infty} \mathrm{E}\big[\big(S_{\text{appr}}^{(n)} - S_{\text{ex}}^{(n)}\big)^2\big] = 0 \qquad \text{and} \qquad (ii')\quad \lim_{n\to\infty} \mathrm{E}\big[\big(S_{\text{ex}}^{(n)} - T^{(n)}\big)^2\big] = 0; \tag{A.38}$$

while $A.37(i)$ is the same as (i'), $A.37(ii)$ is a consequence of (i'), (ii') and the triangle inequality.

Defining the *antirank* of $V_i^{(n)}$ with respect to $U_i^{(n)}$ by $Q_{i;*}^{(n)} := \{r : R_r^{(n)} = i\}$ (so that $R_{Q_{i;*}^{(n)}}^{(n)} = i$), the sequence $\big(Q_{1;*}^{(n)}, \ldots, Q_{n;*}^{(n)}\big)$ is uniformly distributed over $\big\{1, \ldots, n\big\}$ in view of the independence between the $U_i^{(n)}$'s and the $V_i^{(n)}$'s. Reordering terms, we have

$$S_{\text{appr}}^{(n)} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(a_{\text{appr}}^{(n)}(i) - \overline{a}_{\text{appr}}^{(n)}\big) b_{\text{appr}}^{(n)}\big(Q_{i;*}^{(n)}\big) \quad \text{and} \quad S_{\text{ex}}^{(n)} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(a_{\text{ex}}^{(n)}(i) - \bar{\varphi_U}\big) b_{\text{ex}}^{(n)}\big(Q_{i;*}^{(n)}\big).$$

Write $S_{\text{ex}}^{(n)} = S_{*;1}^{(n)} + S_{*;2}^{(n)}$, where

$$S_{*;1}^{(n)} := n^{-1/2} \sum_{i=1}^{n} \big(a_{\text{appr}}^{(n)}(i) - \overline{a}_{\text{appr}}^{(n)}\big) b_{\text{ex}}^{(n)}\big(Q_{i;*}^{(n)}\big) \quad \text{and} \quad S_{*;2}^{(n)} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(b_{\text{ex}}^{(n)}(i) - \bar{\varphi_U}\big)\big(a_{\text{ex}}^{(n)}\big(R_{i;*}^{(n)}\big) - a_{\text{appr}}^{(n)}\big(R_{i;*}^{(n)}\big)\big),$$

where $R_{i;*}^{(n)} := \{r : Q_r^{(n)} = i\}$ denotes the antirank of $U_i^{(n)}$ with respect to $V_i^{(n)}$. Assumption (A5), (A.33), Lemma V.1.6a, and Theorem V.1.6a from Hájek and Šidák (1967) together imply $\lim_{n\to\infty} \mathrm{E}\big[\big(S_{\text{appr}}^{(n)} - S_{*;1}^{(n)}\big)^2\big] = 0$ and $\lim_{n\to\infty} \mathrm{E}\big[\big(S_{*;2}^{(n)}\big)^2\big] = 0$, which, along with the triangle inequality, establishes (i') in (A.38).

Let $\mathbf{U}_{(\cdot)}^{(n)} := \big(U_{(1)}^{(n)}, \ldots, U_{(n)}^{(n)}\big)'$ and $\mathbf{V}_{(\cdot)}^{(n)} := \big(V_{(1)}^{(n)}, \ldots, V_{(n)}^{(n)}\big)'$ denote the order statistics for the n-tuples $\{U_i^{(n)}\}_{i=1}^n$ and $\{V_i^{(n)}\}_{i=1}^n$, respectively. Because the antiranks $R_{1;*}^{(n)}$ are uniformly

distributed and independent of $R_1^{(n)}, \ldots, R_n^{(n)}$, the $R_{1;*}^{(n)}$th order statistic $U_{(R_{1;*}^{(n)})}$ is uniformly distributed over the unit interval (the same is true for the $Q_{1;*}^{(n)}$th order statistic $V_{(Q_{1;*}^{(n)})}$). Write $T^{(n)} = T_{*;1}^{(n)} + T_{*;2}^{(n)}$, where

$$T_{*;1}^{(n)} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(a_{\mathrm{ex}}^{(n)}(i) - \overline{\varphi}_U\big)\varphi_V\big(V_{(Q_{i;*}^{(n)})}\big) \text{ and } T_{*;2}^{(n)} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\varphi_V(V_i) - \overline{\varphi}_V^{(n)}\big)\big(\varphi_U(U_{(R_{i;*}^{(n)})}) - a_{\mathrm{ex}}^{(n)}(R_{i;*}^{(n)})\big).$$

Then (A.33) and Theorem V.1.5a from Hájek and Šidák (1967) imply that

$$\lim_{n\to\infty} \mathrm{E}\big[\big(S_{\mathrm{ex}}^{(n)} - T_{*;1}^{(n)}\big)^2\big] = 0 \qquad \text{and} \qquad \lim_{n\to\infty} \mathrm{E}\big[\big(T_{*;2}^{(n)}\big)^2\big] = 0,$$

which establishes (ii') in (A.38). □

Proof of Proposition 2.2. All expectations in this section are under $\mathrm{P}_{\boldsymbol{\mu},\mathbf{L},f}^{(n)}$, unless otherwise specified; $\mathbf{R}_i^{(n)}$ stands for $\mathbf{R}_i^{(n)}(\mathbf{L})$, $i = 1, \ldots, n$. For part (i) of the proposition to hold, it is sufficient that, for $\underset{\sim}{\mathbf{T}}_{\mathbf{L},f;\mathrm{ex}}^{(n)}$ and $\underset{\sim}{\mathbf{T}}_{\mathbf{L},f}^{(n)}$ in (2.9) and (2.14),

$$\big(\underset{\sim}{\mathbf{T}}_{\mathbf{L},f}^{(n)}\big)_{rs} = \big(\underset{\sim}{\mathbf{T}}_{\mathbf{L},f;\mathrm{ex}}^{(n)}\big)_{rs} + o_{L^2}(1) \quad \text{for all } r, s \in \{1, \ldots, k\}, \text{ as } n \to \infty. \tag{A.39}$$

First, fix $r \neq s \in \{1, \ldots, k\}$. Then,

$$\big(\underset{\sim}{\mathbf{T}}_{\mathbf{L},f;\mathrm{ex}}^{(n)}\big)_{rs} \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathrm{E}\left[J_{f_r}\left(U_{1r}^{(n)}\right) \big| R_{ir}^{(n)}\right] \mathrm{E}\left[F_s^{-1}\left(U_{1s}^{(n)}\right) \big| R_{is}^{(n)}\right]$$

by independence between distinct components, and

$$\big(\underset{\sim}{\mathbf{T}}_{\mathbf{L},f}^{(n)}\big)_{rs} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(J_{f_r}\big(\frac{R_{ir}^{(n)}}{n+1}\big) F_s^{-1}\big(\frac{R_i s^{(n)}}{n+1}\big) - \overline{J_{f_r}}^{(n)} \overline{F_s^{-1}}^{(n)}\Big).$$

Letting $\phi_U = J_{f_r}$ and $\phi_V = F_s^{-1}$, (A.39) (for $r \neq s$) thus directly follows from Lemma A1. For $r = s$, the Hájek projection theorem for linear rank statistics and the convergence rate of Riemann sums imply

$$\begin{aligned}\left(\underset{\sim}{\mathbf{T}}^{(n)}_{\mathbf{L},f;\mathrm{ex}}\right)_{rr} &:= n^{-\frac{1}{2}}\sum_{i=1}^{n}\left(\mathrm{E}\left[J_{f_r}\left(U^{(n)}_{ir}\right)F_r^{-1}\left(U^{(n)}_{ir}\right)\big|R^{(n)}_{ir}\right]-1\right)\\ &= n^{-\frac{1}{2}}\sum_{i=1}^{n}\left(J_{f_r}\left(\frac{R^{(n)}_{ir}}{n+1}\right)F_r^{-1}\left(\frac{R^{(n)}_{ir}}{n+1}\right)-1\right)+o_{L^2}(1)\\ &= n^{-\frac{1}{2}}\left(\frac{1}{n}\sum_{i=1}^{n}J_{f_r}\left(\frac{i}{n+1}\right)F_r^{-1}\left(\frac{i}{n+1}\right)-\int_0^1 J_{f_r}(u)F_r^{-1}(u)\mathrm{d}u\right)+o_{L^2}(1)=o_{L^2}(1)\end{aligned}$$

as $n\to\infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L},f}$. This establishes part (i) of Proposition 2.2. As for part (ii), it follows from the results in Hallin and Werker (2003) that $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{L},\boldsymbol{\mu},f;\mathrm{ex}}=\boldsymbol{\Delta}^{(n)*}_{\mathbf{L},\boldsymbol{\mu},f}+o_{L^2}(1)$ as $n\to\infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L},f}$. This, along with part (i) of the proposition and the triangle inequality, implies part (ii). □

Proof of Proposition 2.3 . In order to establish part (i) of the proposition, it is sufficient to show that, for every $r\neq s\in\{1,\ldots,k\}$, $\left(\underset{\sim}{\mathbf{T}}^{(n)}_{\mathbf{L};f}\right)_{rs}=\left(\mathbf{T}^{\diamond(n)}_{\mathbf{L},\boldsymbol{\mu};f,g}\right)_{rs}+o_{L^2}(1)$ as $n\to\infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L},g}$. Let $\mathbf{V}^{(n)}_i:=\mathbf{G}\left(\mathbf{Z}^{(n)}_i\right)=:(V^{(n)}_{i1},\ldots,V^{(n)}_{ik})'$, $i=1,\ldots,n$. The rank of $V^{(n)}_{ij}$ amongst $V^{(n)}_{1j},\ldots,V^{(n)}_{nj}$ is $R^{(n)}_{ij}(\mathbf{L})$ for each $j=1,\ldots,k$. The claim follows from Lemma A.1 by taking score functions J_{f_r} and F_s^{-1}.

The proof for parts (ii) and (iii) follow from that of Theorem 3.2(ii) and (iii) in Ilmonen and Paindaveine (2011). However, the presence of asymmetry in the independent components implies different cross-information matrices. The result is obtained, via Le Cam's Third Lemma, from an evaluation of the covariance matrix in the asymptotically normal joint distribution of $\boldsymbol{\Delta}^{\diamond(n)}_{\mathbf{L},\boldsymbol{\mu};f,g}$ and (2.8) under $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L},g}$. That covariance matrix follows from the covariance of $\boldsymbol{\Delta}^{\diamond(n)}_{\mathbf{L},\boldsymbol{\mu};f,g}$ and $\boldsymbol{\Delta}^{(n)}_{\mathbf{L},\boldsymbol{\mu},g}$, under $\mathrm{P}^{(n)}_{\boldsymbol{\mu},\mathbf{L},g}$ which depends on

$$\mathrm{E}\big[\mathrm{vec}\big(\mathbf{T}^{(n)}_{\boldsymbol{\mu},\mathbf{L},g}\big)\mathrm{vec}\big(\mathbf{T}^{(n)}_{\boldsymbol{\mu},\mathbf{L},f,g,\diamond}\big)'\big]=\sum_{\substack{r,s,p,q=1\\ r\neq s}}^{k}\mathrm{E}\big[\big(\mathbf{T}^{(n)}_{\boldsymbol{\mu},\mathbf{L},g}\big)_{r,p}\big(\mathbf{T}^{(n)}_{\boldsymbol{\mu},\mathbf{L},f,g,\diamond}\big)_{s,q}\big]\mathbf{e}_p\mathbf{e}_q'\otimes\mathbf{e}_r\mathbf{e}_s'.$$

Evaluating this expression eventually yields the value of $\mathbf{G}_{f,g}$ appearing in (2.15) for the cross-information matrix. □